Automated Image Analysis Of Estrogen Receptor Immunohistochemistry In Breast Cancer

by

©Joel Mercer, B.Eng, M.D.

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and applied Sciences
Memorial University of Newfoundland

May 2014

St. John's, Newfoundland, Canada

Abstract

Breast cancer is among the leading causes of cancer death in women. The estrogen receptor test (ERT) guides therapy and has a direct impact on patient survival. Recent data have shown the intepretation of ERT can vary both within and between laboratories.

The aim of this project was to develop an automated system to classify digital images of ERTs as positive or negative. The slides were classified by two expert breast pathologists prior to being analyzed by the software. Two data sets were used for analysis, one from Eastern Health (Newfoundland) with 60 cases and the other from the UK with 275 cases. The UK dataset was also analyzed by two existing estrogen receptor analysis systems.

The developed algorithm made use of a color deconvolution algorithm to separate the histological stains. It identified the individual cells by use of mathematical morphological operations and a marker based watershed algorithm. Finally, each cell was classified based on its concentrations of stains; if the proportion of positive cells reached a threshold then the case was deemed positive.

Overall, the program reached a sensitivity and specificity of 100% for the first dataset. For the second data set 94% sensitivity and 95% specificity. This was an increase of 10% in sensitivity over the compared software. Given the history of false negatives with the ERT, an increase in sensitivty could offer increased patient safety with use of this software as a quality assurance system. Furthermore, its use could be extended to any immunohistochemical stain of any tissue, human or otherwise.

Acknowledgements

I would like to take this opportunity to thank all the people who have been instrumental in the completion of this thesis. I will start by thanking my supervisors, Dr. R. Gosine and Dr. G. Mann, without their continued support and patience I'm sure this project would never have seen completion. They have been extremely flexible over the years and have worked around my changing schedule. Next, I would like to thank my wonderful fiancee, Dr. Eugenia Khorochkov. Without her help and encouragement this project would likely have remained a stack of papers on my desk. Finally, I would like to thank my parents, Aubrey and Maureen Mercer, for their lifelong support and patience with my seemingly endless schooling. Words cannot express my depth of gratitude to all these people and everyone else that supported me in this endeavor.

Contents

Chapter 1

Introduction

In this chapter I provide background information on breast cancer and how the estrogen receptor test fits into the prognosis as well as how it is performed. Also included will be a review of the literature that is relevant to this work and the contributions made by this thesis.

1.1 Background

This section provides a brief background review of breast cancer taking a top down approach. This review starts with an overview of breast cancer and then going deeper into the types of tests performed and how they affect prognosis.

1.1.1 Breast Cancer

There is no greater feared cancer among women than carcinoma of the breast. According to the American Cancer Society[1], there were 192,370 new cases of invasive breast cancer in 2009, and 40,170 women are expected to die—second only to lung cancer as the highest cause of cancer death in women. Despite advancements in early detection and treatment,

one in four women with breast cancer will die because of it.

There are nine major factors affecting breast cancer prognosis [2]:

Size of Primary Carcinoma: Invasive carcinomas smaller than 1cm have an excellent prognosis in the absense of other signs of poor prognosis

Metastatic Lymph Node Involvement The first place a cancer will spread (metastasize) is to adjacent lymph nodes, ergo the presence of cancer in the adjacent nodes is a poor prognostic predictor. For example, no lymph node involvement has a 5 year survival of 90% whereas 16 positive lymph nodes has a 5 year survival less than 50%.

Distant Metastases: When the cancer has moved beyond the adjacent lymph nodes it is referred to as a distant metastasis. Patients with distant metastases are rarely curable.

Grade of the Carcinoma: Grading of tumors looks at how well differentiated the cells are, that is, how much they look like the normal cells around them. The less differentiated the cells, the worse the prognosis.

Histological Type of the Carcinoma: Certain subtypes of breast carcinoma are associated with a better prognosis.

Estrogren and Progesterone Receptor Status: The presence of hormone receptors indicates a slightly better prognosis. However, the major reason for testing for their presence is to predict the cancer's response to hormone therapy. The highest rate of response (˜80%) to anti-estrogen therapy is seen when both estrogen receptors and progesterone receptors are present.

Aneuploidy: Carcinomas with abnormal DNA have a worse prognosis than carcinomas with DNA similiar to normal cells.

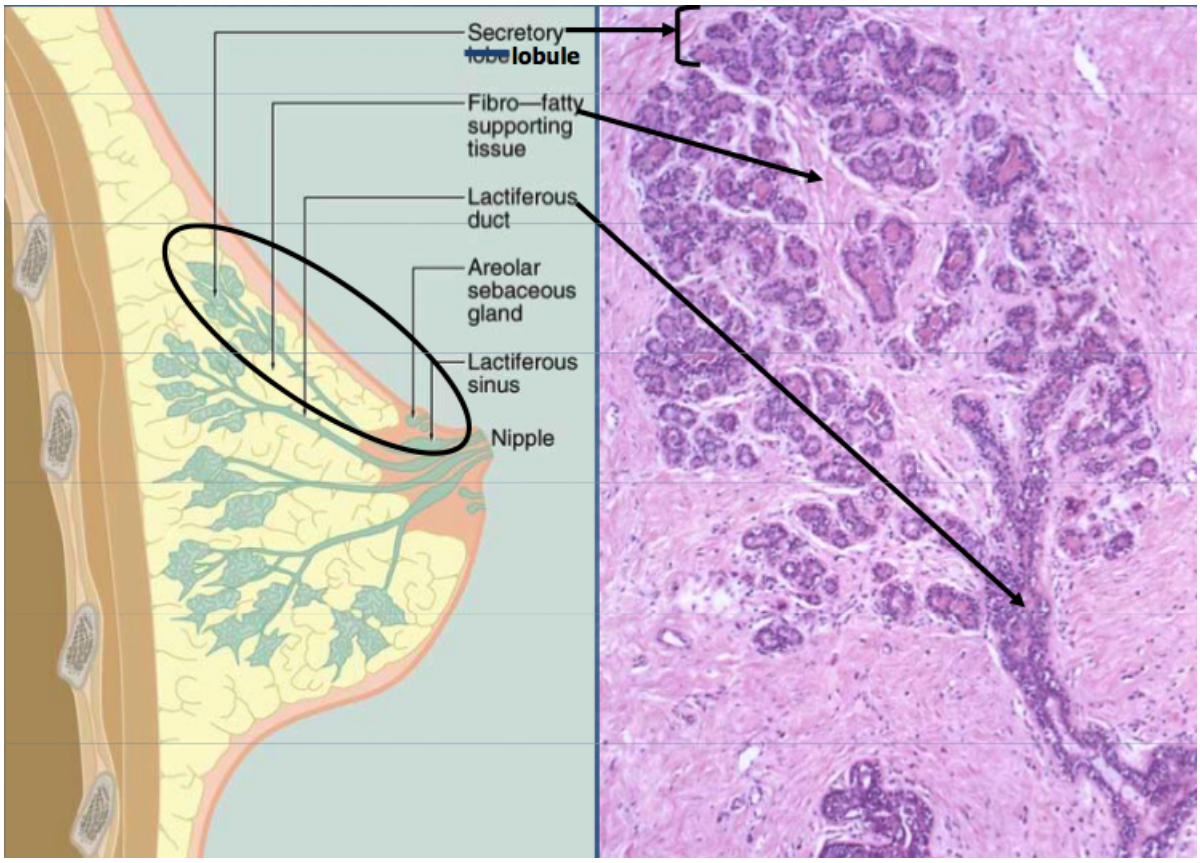

Figure 1.1: Histological Section of Human Breast [3]

Overexpression of HER2/NEU: Over-expression of this membrane-bound protein is almost always caused by a carcinogenic mutation and is associated with a poor prognosis.

This thesis will focus on the presence or absence of estrogen receptors as determined via immunohistochemical staining, which will be explained in the following sections.

1.1.2 Histology

Histology is the study of structure and arrangement of cells within tissues at a microscopic level. It is typically performed by taking slices (or sections) of a tissue, staining the tissue to highlight certain features, and looking at it under a microscope.

When visualizing a histological slide it is important to remember it is a two-dimensional

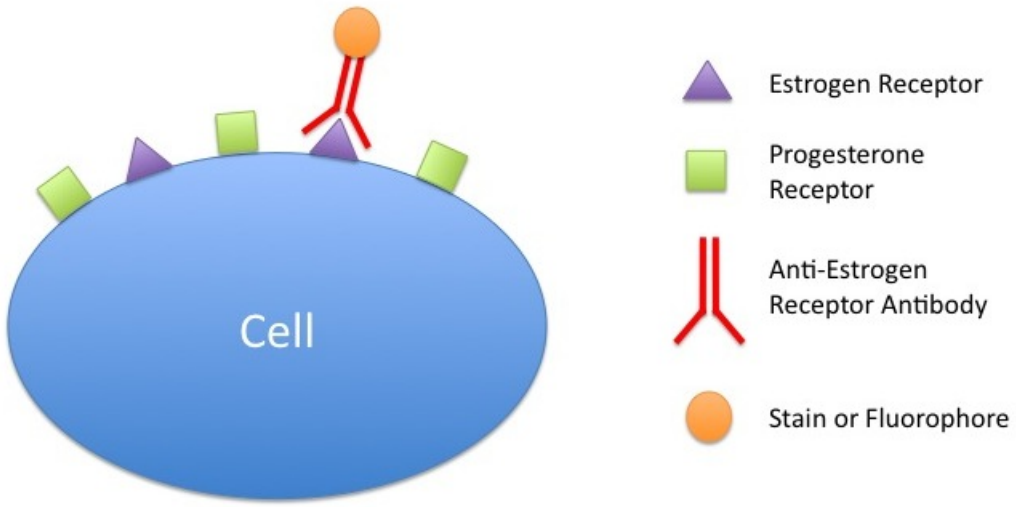

Figure 1.2: Immunohistochemistry

sampling of a three-dimensional structure. It is taking a slice through the anatomy and looking at the microscopic features of the cells. This is shown in figure 1.1 where three-dimensional structures of the breast are matched with histological features in the stained slide.

Many diseases manifest at the microscopic level long before any visible changes to the tissue may occur, so histology is a powerful tool in formulating a diagnosis or prognosis of a disease.

1.1.3 Immunohistochemistry

Immunohistochemistry refers to the process of localizing proteins within the cells of a tissue section. This is accomplished by exploiting the principle of antibodies binding to specific antigens in biological tissue. Assuming an appropriate antibody is available, the binding can be visualized by tagging the antibody with a stain or fluorophore (a fluorescent chemical compound which can emit light upon excitation). When the antibody binds to the antigen it will be visible under a microscope. Immunohistochemical staining is widely used in the

analysis of abnormal cells such as those found in cancerous tumors: the presence or absense of certain proteins can confirm a diagnosis or give a more definite prognosis.

Immunohistochemistry is of great advantage to biomedical studies because it allows antigen specific analysis while maintaining tissue morphology—i.e. the architecture of the cells within the tissue. It is difficult, however, to compare results between labratories because of different methods of staining and analysis.

Estrogen Receptor Test

The most important prognostic determinants for most cancers are the histological type of the tumor, the histological grade of the tumor, tumor size and the presence of lymph node metastases.

An additional important test for breast cancer is the *estrogen receptor* test: a test designed to ascertain the proliferation of estrogen receptors in breast carcinoma and, consequently, predict the response to hormonal therapies.

Estrogen receptors are localized to the nuclei of epithelial cells. Since estrogen receptors are confined to the nucleus, an immunohistochemical stain of this type would be referred to as a *nuclear stain*. If the protein in question was also present in the cytoplasm it would be referred to as a *cytoplasmic stain*. This work will primarily address nuclear staining.

About 7% of normal breast epithelial cells have estrogen receptors [4]; this can change if the cells become cancerous. A larger concentration of estrogen receptors—an ER positive patient— is indicative of a better prognosis and a different course of treatment than that of a patient with a normal level of estrogen receptors. The estrogen receptor positive tumors will grow when exposed to estrogen. The current standard of care is to treat these patients with *tamoxifen* which blocks estrogen from binding with the cancerous cell's estrogen receptors. This will often slow down, or even stop, the tumor progression.

Unfortunately, the estrogen receptor test is not as objective as one might hope. Studies

have shown variation between pathologists at different labs [6]. This is not really suprising given the hundreds of cells the pathologist must assess in a single tumor and to make an estimate of the proportion that display estrogen receptors.

1.1.4 Tissue Microarrays

Tissue microarrays [5] (TMAs) are paraffin blocks containing thousands of tissue samples from different patients. They were originally developed to overcome the limitations associated with studying biomarkers using traditional methods. As discussed in the previous section, each tissue sample would need to have a histological section cut, individually stained and then embedded in paraffin wax. So in order to study a new stain or biomarker hundreds of such slides would have to be prepared.

Enter tissue microarrays, a technique that allowed the ability to assay hundreds or even thousands of patient tissues on a single slide. Each sample is represented as a small (0.6-2.0 mm diameter) histological section (a "core") arranged in a grid pattern—this allows easy linkage to clinical information using a simple coordinate system.

There are several key advantages to using tissue microarray technologies. Firstly, they allow greater use of limited tissue resources since each TMA only uses a small section from the original sample. Secondly, TMAs result in a significant cost savings in terms of the amount of reagent and technician time required to stain one instead of hundreds of slides. Thirdly, TMAs allow much larger studies to be completed—instead of dozens of samples, a researcher could use hundreds. This would have a substantial impact on the quality, significance and reliability of biomarker studies, especially in cases where the biomaker was only present in a small percentage of cancers. Finally, since all the cores are stained together there is a guarantee of consistency between cores. They would have been stained using the same stain concentration, the same amount of fixation time, same antibodies etc. Therefore, studies done with TMAs will not suffer from problems of variable staining levels.

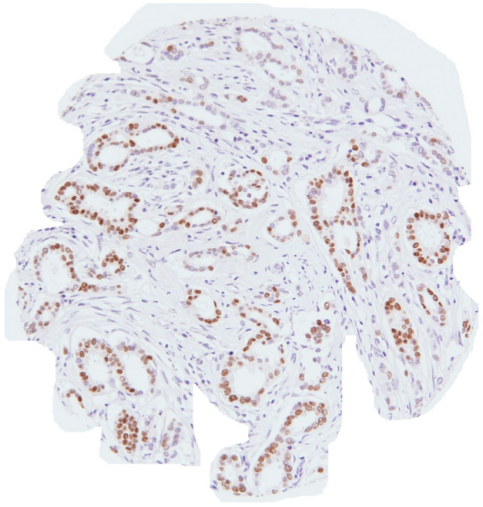

Figure 1.3: Estrogen Receptor Test

While there are considerable advantages in using tissue microarrays, there are also drawbacks. TMA's suffer from many of the same problems as traditional methods. Specifically they are dependent on good tissue quality, appropriately validated antibodies and standardized laboratory techniques. Addditionally, the technical skills required to produce TMA's are substantially greater than those required to embed traditional blocks and, therefore, are not available in many smaller centers.

1.2 Problem Statement

The problem being addressed in this thesis is the automated image analysis of the estrogen receptor test; that is, a digitalized slide will be analyzed by a custom software application and classifed as either positive or negative. As discussed in previous sections, the estrogen receptor test is a prognostic test done for breast cancer patients to determine their responsiveness to hormone blocking therapy. The pathologist is required to determine the percentage of epithelial cells that are positive for estrogen receptors. Based on this analysis it will be

decided if the patient will benefit from hormone blocking therapy—it is important to note that all other therapies will remain the same regardless of the outcome of this test (surgery, radiation, chemotherapy etc). This test has often been subject to intense media scrutiny [7] due to aforementioned interlaboratory varation and lack of objective standardization.

A sample of a stained slide can be seen in figure 1.3. This test is ideal for automation as it is known for its subjectivity and determines whether or not a patient receives potentially life-saving therapy. As digital slide scanners become more common in pathology laboratories, the automation of certain tests is sure to follow. The aim of this work is to develop algorithms to determine the percentage of estrogen receptor positive cells in a stained histological section in an objective, reproducible manner.

At a high level this involves 3 different stages:

Cell Segmentation: The first stage attempts to separate the nucleus of the cells from the cytoplasm and other background artifacts.

Color Separation: The second stage attempts to determine the quantity of each stain present in each cell.

Cell Classification: Based on data obtained in the second stage the cell is classified as either ER positive or ER negative.

Each of these steps will be explored in much greater detail in the coming chapters.

1.3 Contributions

This thesis will be the first to study to combine color deconvolution techniques with mathematical-morphological operators for the evaluation for immunohistochemical stains. Both of these techniques are well documented and are certainly not new to biomedical imaging applications.

It is difficult to assess how well this algorithm will perform against others that have been developed for immunohistochemical stains as there is no common dataset that is used as a standard by all researchers. Furthermore, most of the software developed for this purpose is commericially developed, often carrying a large price tag as well as custom hardware. The algorithm developed here will be compared against one commercial software and one freely-available software designed to evaluate estrogen receptor tests.

Chapter 2

Related Work

This chapter provides a review of the literature related to the automated scoring of histological images. This includes work done on the segmentation of cells, the separation of immunohistochemical stains as well as complete systems designed to perform both tasks.

Using image analysis on estrogen receptor tests is not a new idea. Bacus et al. [8] evaluated computer-assisted image analysis of estrogen receptor tests in the late 1980's. They used bandpass filters to separate the stain colors and then measured the intensity of the corresponding greyscale image. Their results showed a correlation between the measured intensity values and the pathologist's visual score of the estrogen receptor test. At this point, however, the entire process was heavily supervised by a trained pathologist and required more work than having the pathologist simply score the slide.

Since that time many approaches have been proposed requiring differing levels of supervision and direction by a pathologist. Individual components of the problem have been tackled; i.e. cell segmentation and color separation. In some cases full end-to-end systems have been developed to allow direct comparison with pathologist scoring and patient outcomes.

Most of the color separation techniques involve changing the color space and using one or more color channels, to classify the staining of individual cells. Some systems are based on

an HSI (Hue, Saturation, Intensity) classification [9] [10] [19] [10] while some use a LUV [12], CMYK [13] or even YIQ [14].

Willemse et al [15] investigate using fixed threshold RGB and interactive threshold greyscale in the segmentation of immunohistochemical images. This was an early study, done in 1994, that showed RGB had a definite advantage over greyscale when analyzing IHC images. All subsequent work would focus on color image analysis.

Ruifrok and Johnston [16] created the color deconvolution algorithm—described in greater detail in chapter 3—which is intended to separate mixtures of inks and dyes into their base components. This type of algorithm is ideal for image processing in histology because the stains can be individually quantified with color deconvolution. Several years later Ruifrok et al [17] showed color deconvolution to be significantly better at quantifying histochemical staining as compared to HSI quantification.

Brey et al [18] compared previously published color separation techniques in the quantification of DAB[1] stained soft tissue from rabbits. Included in this comparison are color deconvolution, HSI, G/B (RGB green intensity/blue intensity), Brown (brown = blue - 0.3* RGB), Blue (RGB blue channel), Hue (from HSI), Green (RGB green channel) and the author's own normalized blue (BN)[2]. The approaches were assessed based on their level of agreement of individual pixel staining with two experts. The best approach was found to be the author's own BN approach, closely followed by color deconvolution. The worst performance was found to be based on the hue channel of HSI, which is interesting because this is the most commonly used approach in quantifying stain intensities.

Similarly, Pham et al. [13] compared CMYK, RGB, nRGB (normalized RGB) and HSL color models on their ability to separate hematoxylin from DAB staining. They found that the Yellow channel of the CMYK model had the greatest power to differentiate between the

[1]Diaminobenzidene a brown IHC stain

[2]RGB blue channel normalized by red, green and blue values

two stains. This makes intuitive sense as the CMYK model is based on a subtractive color system, whereas the rest use an additive color system.

Lehr et al [11][19] used the commerical software Adobe Photoshop to separate immunohistochemical stains. They showed that the greyscale quantification of cell nuclei correlates to biochemical tests—the gold standard for estrogen receptor quantification. Their approach used Photoshop's built in "wand tool" and histogram function to segment and then quantify the staining intensity. Also covered was the separation of two stains using Photoshop functions as compared with using a band-pass filter. They showed that using the HSI color system, with the aforementioned approach, they were able to produce better results than using band pass filters. While this approach does yield good results, it places too large a burden on the user as it requires them to manually segment and select color ranges for each image.

Almost all of the papers on segmentation have used some sort of marker-based watershed algorithm, whether a direct watershed or one using more complex criteria for the selection of minima. Several researchers attempted other popular segmentation techniques, but they were found to be inappropriate for segmenting clusters of cells.

Malpica et al. [20] applied a marker-based watershed algorithm to bone marrow and blood samples. They achieved good segmentation results using the morphological operation *h-dome*—this served to reduce oversegmentation of the cell clusters. This approach was only applied to bone marrow or peripheral blood samples.

The watershed algorithm was not the only segmentation algorithm used in pathology images; Bamford et al. [21] used active contours to segment cells in Papanicolaou-stained cervical cell images (Pap smears). This approach yielded a more accurate segmentation than watershed or other morphological transforms, but it could not separate clusters of cells, which is fundamental in analyzing the estrogen receptor tests. So while this approach yields a more accurate delineation of the nucleus it could not be used for segmentation in the ER

test.

Mao et al [22] applied a marker-based watershed algorithm to p53 immunohistochemistry in bladder cancer. This test poses simliar challenges to the estrogen receptor staining in breast cancer. They compared a basic grayscale watershed with a complemented distance transform and finally their own marker-based watershed algorithm. They proposed that markers (local minima) could be merged based on the topological surface between the two points; if there was a large transition between the two they were likely different nuclei and if a small transition exists they probably represented an oversegmentation of a single nucleus. They show that this approach resulted in far less segmentation errors than either applying a watershed transform directly to the greyscale image or the complemented distance transform.

Schüpp et al. [23] also used a marker-based watershed algorithm to segment breast cancer nuclei. This paper points to the applicability of mathematical morphology to the detection and segmentation of breast cancer nuclei, despite the relative paucity of papers written on the subject. The authors employed a relatively simple segmentation technique: it blurred the image, applied a threshold to binarize, then applied the mathematical operator known as *ultimate erosion*. This transform found the regional maxima of the complemented euclidean distance map. This was followed by a watershed operation. Overall, this was a fairly standard approach, but showed the strength of mathematical morphology in dealing with cell segmentation.

Latson et al. [24] attempted to use a fuzzy clustering algorithm to segment breast cancer slides stained with hematoxylin and eosin. This approach was used in conjunction with a watershed algorithm to separate groups of nuclei. In the end, the fuzzy clustering approach was dropped in favor of a fixed threshold in the HSV color space. They found that while up to 25% of nuclei were incorrectly segmented they had in fact missed very few cells. These false segmentations were attributed to nuclear stippling [3] as well as having no algorithm to

[3]Stippling refers to the white-spotted appearance of cell nuclei that results from increased chromatin in

dealng with clusters of nuclei that were not separated by their watershed algorithm.

While some researchers looked at isolated problems (i.e. cell segmentation or color separation), others looked at the end-to-end problem: starting with an immunohistochemically (IHC) stained slide and ending with a proportion of positive cells. After reviewing the literature one notices most systems are developed in isolation; they all demonstrate good results but without offering comparison to previous literature/algorithms, and thus we are left with islands of research. Consequently it becomes difficult to define the state-of-the-art. The following provides an overview of previous systems adding any major pros or cons of each approach.

Schnorrenberg et al [14] developed a system that employed neural networks to help identify and classify breast cancer nuclei. The system used the Y channel from the YIQ color space—the same channel used in black and white televisions—to determine the staining intensity. To locate the nuclei, blocks of the image underwent Single Value Decomposition (SVD) to form a feature vector, which was then fed into the neural network for classification. One major limitation of this paper was that it used images with minimal overlap displayed between nuclear boundaries. This seriously limits the applicability of the system as most real-world slides would have overlapping cells.

Kostopoulos et al [25] developed custom ER scoring software using a black and white texture recognition approach. They used 16 black and white texture features and evaluated 22 cases using a k-nearest neighbours classifer. They achieved an overall accuracy of 86.4%. The system was shown to be superior to an approach using Adobe Photoshop.

Kayser et al [10] developed a web-based system to analyze a variety of different histological slides. It employed a HSI color system to separate the colors and a basic thresholding technique to segment the cells. This paper was different in that it focused more on the useability of their software, as well as its functions, layout and overall architecture. It

the nuclei

raised the point that while commercially available image analyzing systems are sold, they are applied mainly to research and rarely for routine practice. This is probably due to prohibitive costs and general useability issues, but given progressive improvement of the available systems it is likely they will find their way to clinical practice.

Tuominen et al [26] also developed a freely available web-based application for the analysis of estrogen and progesterone receptors entitled *ImmunoRatio*. Their software employed Ruifrok's color deconvolution algorithm for color separation and a combination of adaptive thresholding and watershed algorithm for segmentation. While this seemed to be a well developed system, the authors did not offer much validation of their results. Although the authors stated an impressive correlation with the "Gold Standard" ($r = 0.98$), they fail to identify how it was obtained. Finally, similiar to most of the other systems developed, no other algorithms or software were compared.

Gokhale et al [27] evaluated two commercial systems, ARIOL's SL-50 and Chromavision's Automated Cellular Imaging System against the manual scores of four independent breast pathologists. They used 64 breast cancer specimens stained for estrogen receptors. This was one of the few articles to compare more than one commercial system; so we are able to see both how these systems compare to one another using the expert breast pathologists' as the gold standard. Unfortunately, no information is given regarding the inner workings of the commercial systems but this is to be expected with any commerically available software. Overall, the results were very positive with the systems being functionally equivalent with a 95% concordance with the 4 pathologists.

Nassar et al [28] evaluauted Aperio's new commercially available software for estrogen receptor scoring. They completed a multi-center study involving 260 specimens with three pathologists at each site independently scoring them. Their study showed high concordance between manual microscopy and image analysis and the authors concluded that they were substantially equivalent. While otherwise a well designed study, the authors failed to compare

it with other commerically or freely available software.

Krecsak et al [29] evaluated the utility of the custom developed NuclearQuant v1.13 for evaluation of ER/PR immunohistochemical stains. They do not go into detail on the algorithm as they have made it commericially available, but do state that they use a color deconvolution algorithm in addition to density and contrast measures for segmentation. They employed a relatively small sample size of 16 cases and compare it with an expert breast pathologist. They received excellent concordance with a Cohen's Kappa of 0.859. However, a more thorough discussion of methodology or comparison to other commercially or freely available systems would be good to determine if this algorithm is truly a step forward.

Vrekoussis et al [30] proposed a simple system using the freely available ImageJ software (Developed by NIH) and a special red stain so that an RGB color model could be used. They received promising results but it is of limited practicality given the use of the special stain. Most pathology labs use DAB as their IHC stain of choice. Similiarly, Prasad et al [31] described their custom image analysis software (TissueQuant) and compared it with manual pathologist scoring of 60 breast cancer cases stained for ER/PR. The system used an HSI color system to segment and classify stained cells. They reported a sensitivity of `100%` and specificity of `82.3%`. Unfortunately no direct comparison was offered to existing systems and a specificity of `82.3%` is lower than most published results.

The present work aims to develop a complete system using a novel segmentation technique based on mathematical morpholgy and color deconvolution for stain quantification. Although both commercial software and hardware are available to analyze estrogen receptor tests, they are limited by high costs, manual segmentation by pathologist and lack of data comparing such systems.

Commercial software and hardware exist that performs this function, but typically it must be trained and supervised by a qualified pathologist. Furthermore, the algorithms for these programs are not made available in the public domain for scrutiny. This system will be

evaluated against one freely available and one commericial software, using the consensus of two breast pathologists as the gold standard. Finally, the algorithm will be made available in the public domain for scrutiny, further development and future research.

Chapter 3

Cell Segmentation

As discussed in the literature review chapter, cell segmentation in histopathology has been an active area of research for the past twenty years. Most of the newer techniques focus on segmentation of cells in histological sections that have been stained with hemotoxylin and eosin (H&E), which are the most common histological stains and can be seen in figure 3.1. In these images all the nuclei are stained with hemotoxylin and everything else is counterstained with eosin. This is quite different from the immunohistochemistry (IHC) images in which only the nuclei that are positive for some cellular marker (in our case estrogen receptors) are stained with DAB, and the rest are stained with hemotoxylin (IHC counterstain). As a result, the negative nuclei blend quite well into the background, making their segmentation considerably more challenging. An example of an estrogen receptor negative image can be seen in figure 3.1.

The hemotoxylin and eosin sections are generally used to diagnose pathology. For example, breast cancer would be diagnosed with an H&E stain, whereas an IHC stain would be used to further characterize the breast cancer—i.e. what drugs will the cancer respond to? For diagnostic purposes the geometrical arrangement of the cells becomes quite important. Normal, healthy cells will arrange themselves in a very characteristic manner, any devia-

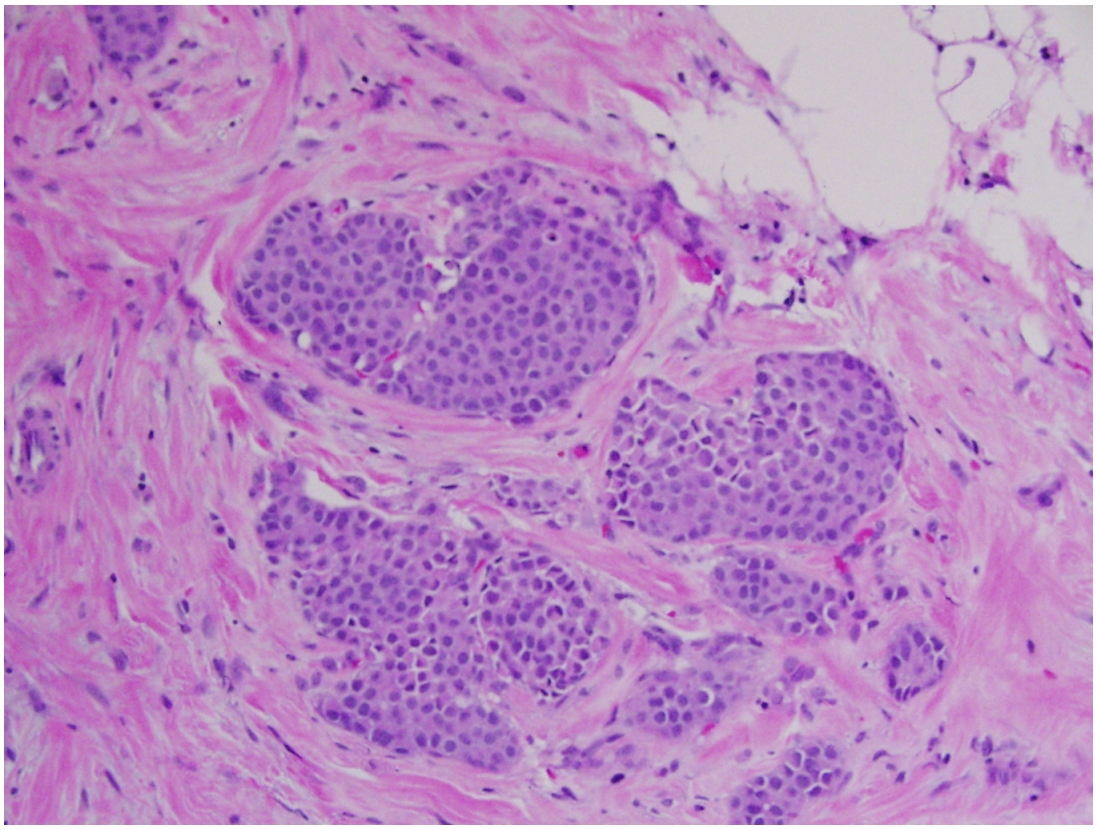

Figure 3.1: Hemotoxylin and Eosin stained section

Figure 3.2: Segmentation Model

tion from this can indicate pathology. Hence, for the purposes of automated diagnosis of histopathological images a very robust segmentation algorithm is needed. Fortunately, we are only concerned about whether a significant proportion of the nuclei are receptor positive and not the geometrical distribution of the nuclei. Therefore, while IHC offers more challenging segmentation, it is not as important to be as accurate.

This chapter begins by outlining the algorithms used in segmenting the cell nuclei. Most of these techniques are based on mathematical morphological operators. An introduction to mathematical morphology is provided in appendix A, while a complete discussion is available in [32].

3.1 Morphology and Cell Segmentation

Any image segmentation system can be divided into pre-processing, segmentation and post-processing as seen in figure 3.2. The following sections will work through an example using the developed segmentation system broken down into the aforementioned components.

3.1.1 Pre-Processing

Pre-processing aims to clean up the image in order to get better segmentation results. The segmentation phase requires a binary image, so the pre-processing stage must convert the color ER/PR slide to a binary image while preserving as much of the cells' shapes as posible.

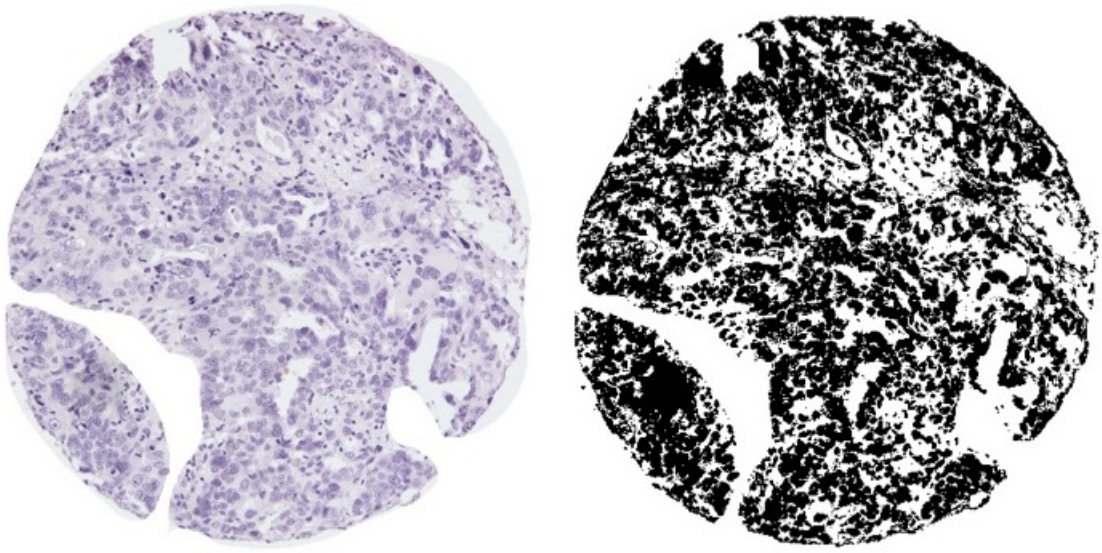

Figure 3.3: Noisy ER slide

The simplest possible pre-processing would be the conversion of the color image to grayscale (the Y channel of the YIQ color system), and then use Otsu's [33] method to threshold the image to result in a binary image. This process does work in some cases, while other cases with a stronger background staining do not work as well. An example of this can be seen in figure 3.3.

However, the primary problem with this approach is that estrogen negative cells, which are stained the same color as the background, tend to get segmented as a part of the background. This would tend to skew the results towards being positive.

The first step in pre-processing the application of a histogram equalization technique to maximize the contrast between the cells and the background. A contrast limited adaptive histogram equalization was used, which was available through the MATLAB image processing toolbox. This algorithm differs from a normal histogram equalization by using multiple histogram equalizations on smaller regions rather than the entire image, as this provides an increase in contrast throughout the image.

This increase in contrast is important for subsequent steps, as it makes it easier to

separate the nuclei from everything else in the image. An example of this can be seen in figure 3.4 with the image before and after equalization being shown. The corresponding pre and post equalization histograms are seen in figure 3.5.

After the histogram equalization, morphological operators are applied to the image to remove some of the background and highlight objects measuring close to a given size. Specifically, we perform a black top-hat by reconstruction (see Appendix on morphological operators). By using a flat disk-shaped structuring element, we can find objects that are round and darker than their surroundings. The choice of size of the structuring element depends on the size of the expected cells, and should be set based on what the smallest cell would measure. In our case, we can measure the size (in pixels) of a lymphocyte which will appear as a small, darkly staining nucleus; this should ensure all cell nuclei are picked up by the segmentation algorithm. The image in figure 3.4 has a top-hat operator applied to it in figure 3.6.

After the top-hat operator has been used, an opening by reconstruction is done to cut down on the remaining noise and the variation within the nuclei that can occur as a result of the state of the chromatin. The result of this subsequent transformation can be seen in figure 3.7.

After this denoising step is applied, Otsu's method [33] is applied to binarize the image in preparation for segmentation. The resultant binarized image can be seen in figure 3.8.

3.1.2 Segmentation

Once we have the binarized images we need to identify which blobs represent cells, which are noise, and which are overlapping cells. We approach the last of these problems first by employing a segmentation algorithm.

After trying several different segmentation methods, a modified watershed algorithm seemed to perform the best. Most attempts at using a watershed [34] algorithm resulted

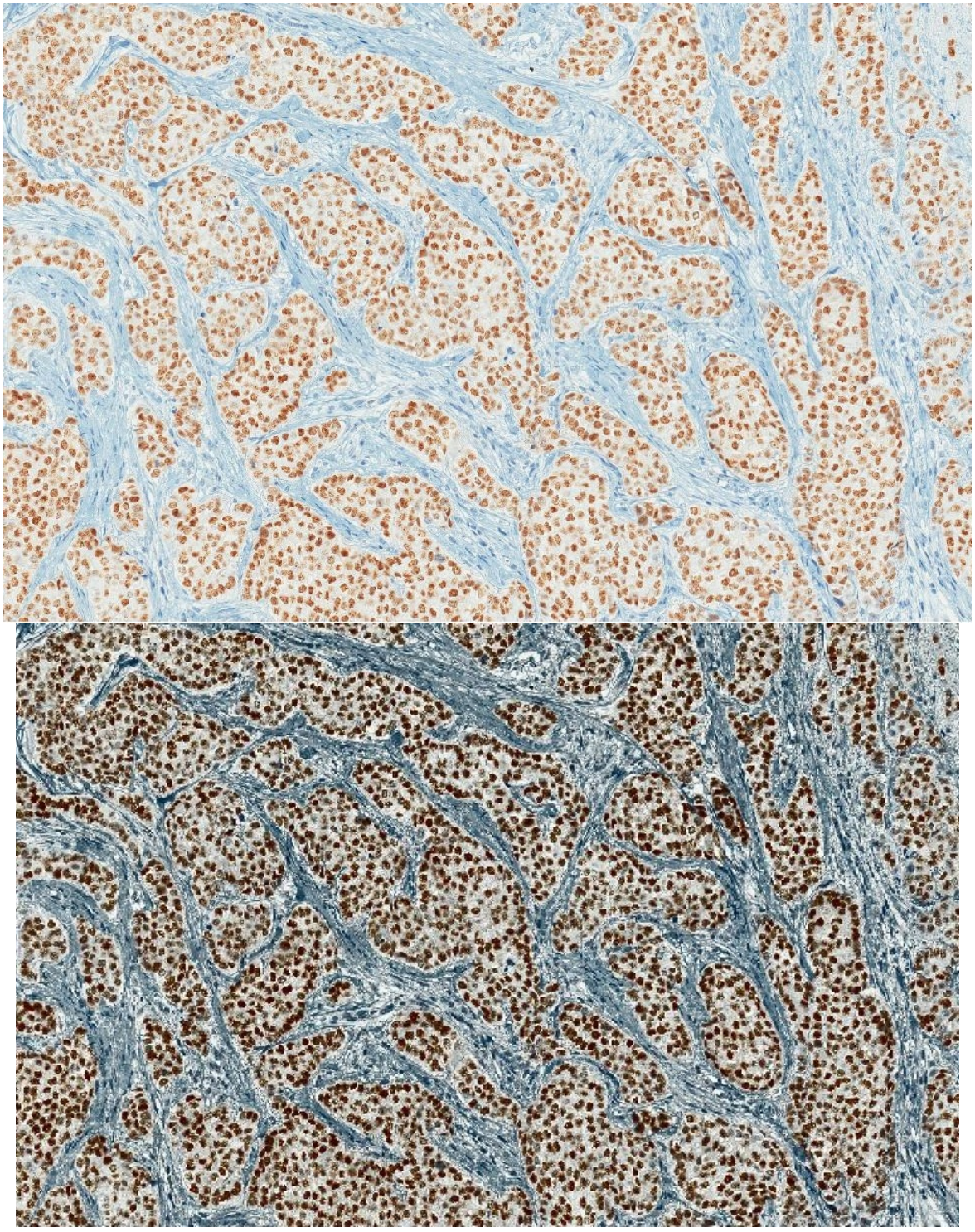

Figure 3.4: Histogram Equalization Algorithm
Top → Original
Bottom → Post Equalization

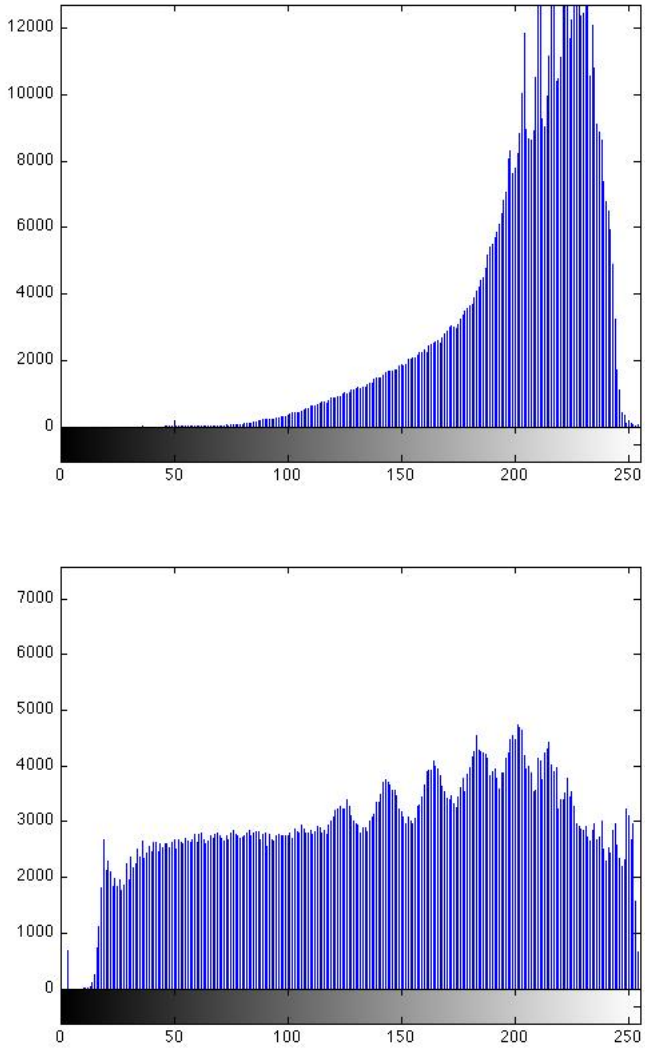

Figure 3.5: Histogram Equalization - Pre & Post Histograms

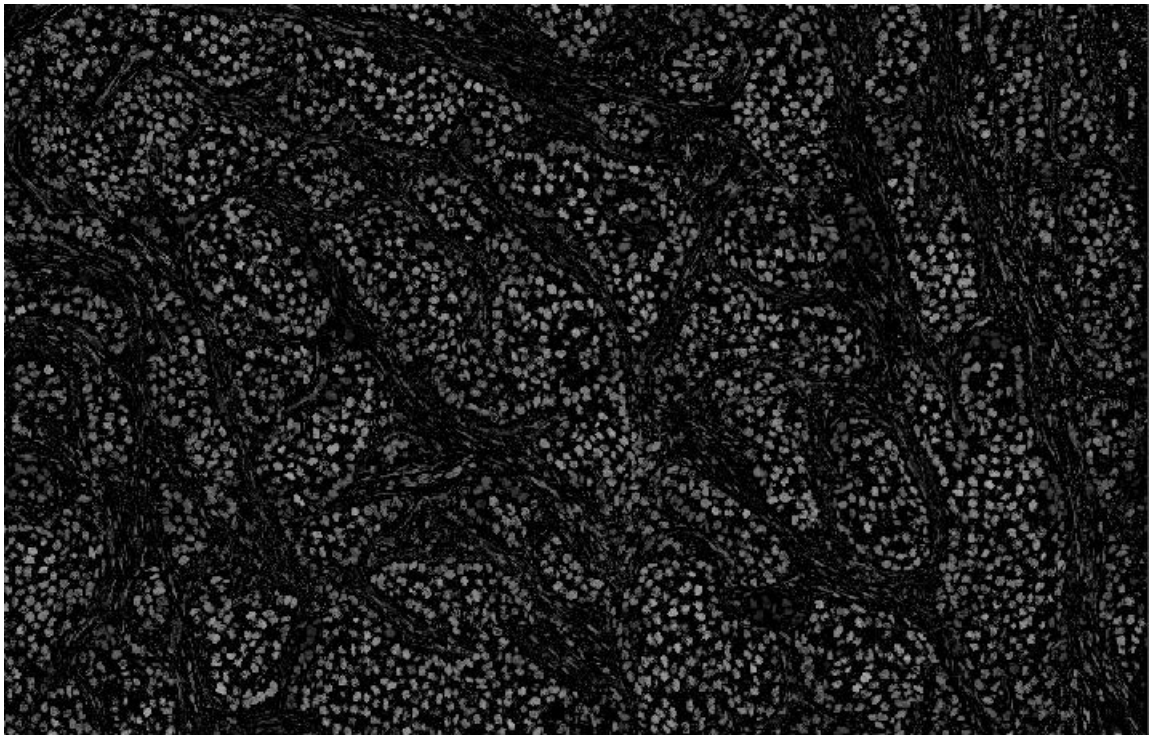

Figure 3.6: Tophat operation applied to sample image

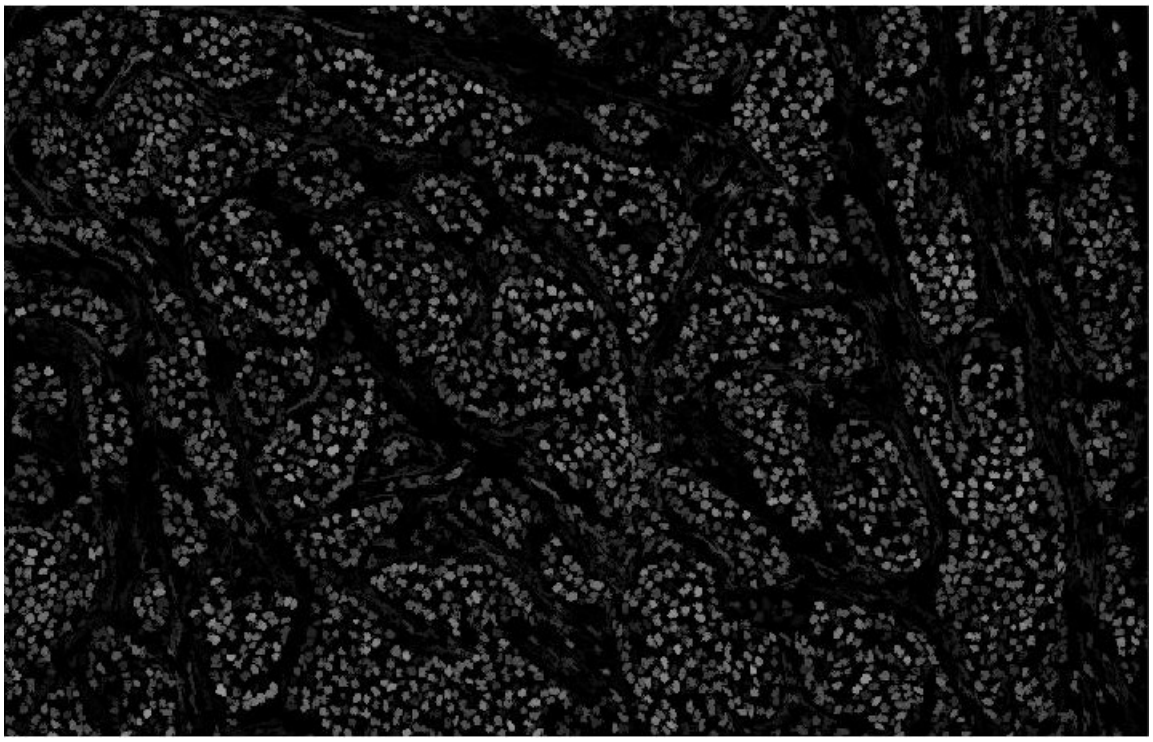

Figure 3.7: Opening by Reconstruction

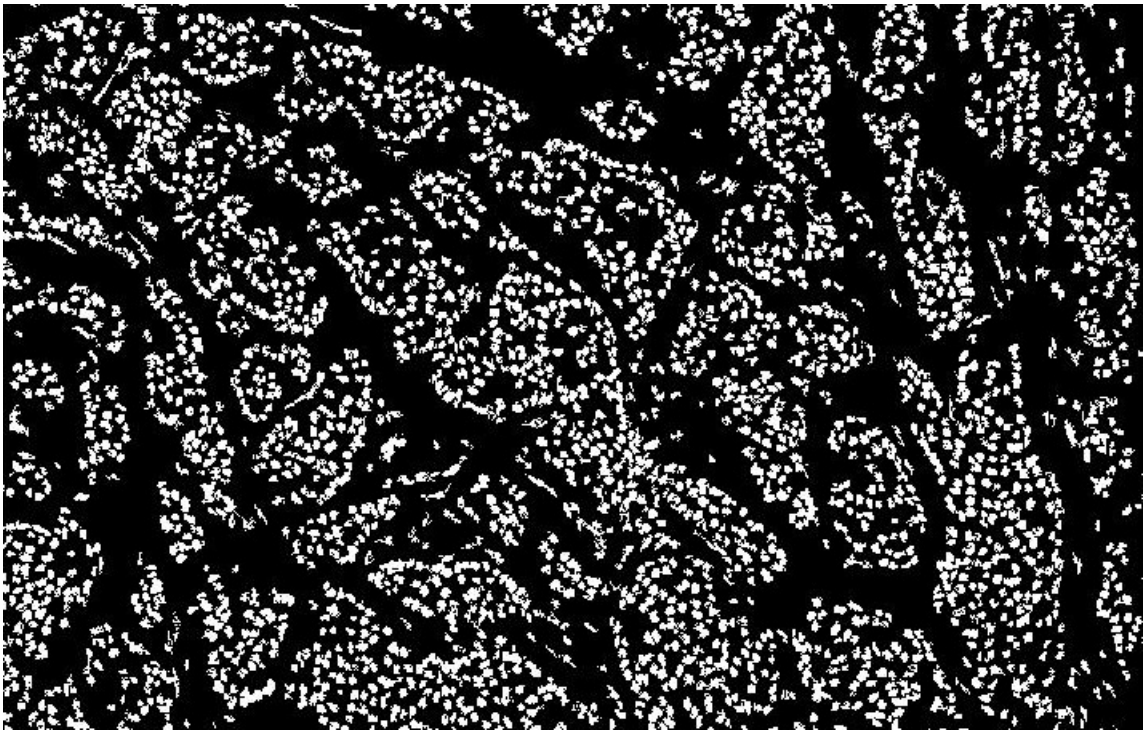

Figure 3.8: Binarization using Otsu's Method

in massive oversegmentation. An example of this was when the watershed algorithm was applied directly to a grey-scale image (whether denoised or not). Somewhat better than this approach was to use a euclidean distance transform on a binary image, but this was still plagued with oversegmentation. Various region-merging techniques were attempted to overcome this, but it added a considerable amount of complexity and computing time while the results were still not significantly better.

The euclidean distance transform assigns a number to each pixel which represents its euclidean distance from the nearest background pixel. The resulting regional maxima are termed *markers*, and when we apply a watershed algorithm it is deemed a marker-based watershed algorithm. This approach can result in oversegmentation when there are small errors in segmenation of the cell boundary; this results in more than one regional maxima per cell, ergo the cell will be split into multiple pieces by the watershed algorithm.

Finally, a technique suggested by Vincent [35] where the result of the distance transform on the binary image was reconstructed using the following equation:

$$Markers = Reconstruction(Dist(I) - 1, Dist(I)) \tag{3.1}$$

Where I represents the binary image and Dist(I) represents the euclidean distance transform. This joins regional maxima which are close together in the image, thereby creating more accurate markers for the watershed algorithm. This approach worked quite well and its inherent simplicity improved performance (both accuracy of segmentation and computation time).

3.1.3 Post-Processing

During the post-processing phase each segmented region is examined individually and based on its size and shape it is either included in further analysis or excluded. Cells of interest

should all be around the same size and have a relatively round shape. The shape criterion can be mathmatically expressed as:

$$Shape - Criterion = \frac{4 * \pi * area}{perimeter^2} \tag{3.2}$$

This function is maximal for a circle where it is equal to 1. The further the shape deviates from a circle the further from one the function will return. Based on experimental results a final value of 0.6 was chosen as a good cutoff to identify properly segmented cells.

The size criteria must be adjusted based on the image set in question. The approach that was used for this project was to measure the area of several normal cells, average the results, and accept fifty percent above and below this number as the upper and lower limits.

The modified watershed algorithm and the cellular inclusion criteria were applied to the binary image seen in figure 3.8. The segmented cells that met the inclusion criteria are shown in figure 3.9 marked in red, superimposed on the original image.

Cells that meet the inclusion criteria can be removed from the image by replacing them with white pixels and the whole segmentation process can start again from the beginning. By approaching this iteratively more cells are identified. However, this approach significantly increases processing time.

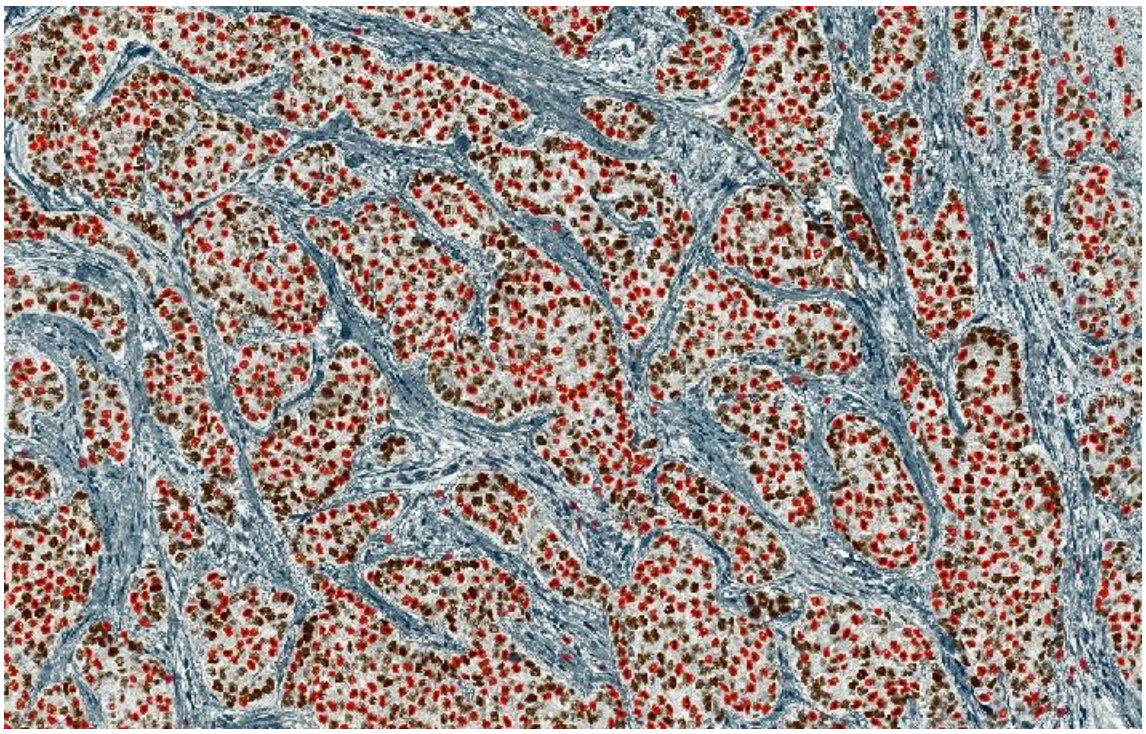

Figure 3.9: Segmented cells (Red) superimposed on original image

Chapter 4

Color Extraction

The first stage in automated immunohistochemical quantification is the extraction of color information. Typically, an immunohistochemical stain will have two components:

Main Stain The stain used to identify the antigen in question

Counter Stain The stain used to mark background tissue

In the case of estrogen receptor tests, *diaminobenzidene* or *DAB* is used to give estrogen receptors a brown coloration and the counter stain hematoxylin gives the background tissue a blue coloration. Visually, an observer can differentiate between the two colors when the antigen staining is high; however, low levels of stained antigen are difficult to separate from areas of dark counter stain. This difficulty in differentiating between these areas of staining leads to variations in the results of manually analyzed staining patterns. Imaging methods that reduce dependence on observer analysis would decrease the time of analysis and lessen the variation in results amoung different labratories.

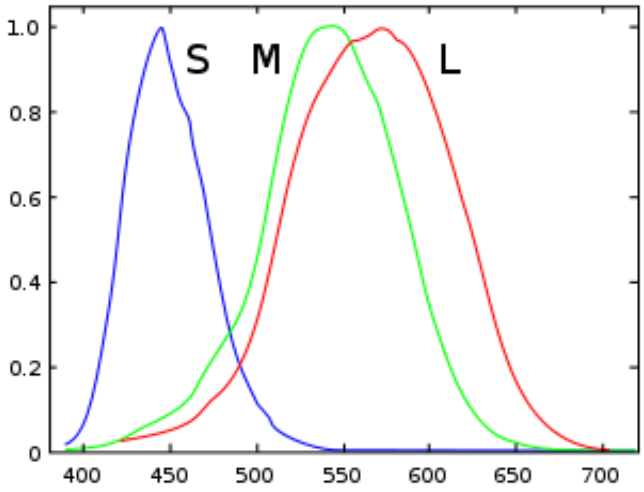

Figure 4.1: Normalized Human Cone Cell Responses

4.1 A Review of Color

Before diving into the techniques used for color separation, a brief review of the principles of color and how they apply to the human eye is warranted. Understanding the mechanics of color and the human eye is especially valuable for this particular problem as we are trying to duplicate human visual analysis with image processing techniques.

4.1.1 Perception of Color in the Eye

The ability of the human eye to distinguish colors is based upon the varying sensitivity of select cells in the retina to light of different wavelengths. The retina contains three types of color-receptor cells, or *cones*.

The first type of cone is most sensitive to light we perceive as violet—or having wavelengths centered around 420 nm. These are sometimes called S-Cones, short-wavelength cones, or blue cones. The second type is most sensitive to light around 530 nm, which we perceive as green. These cones are sometimes called middle-wavelength cones, M-Cones, or green cones. The final type is sensitive to light around 560 nm, which we perceive as red.

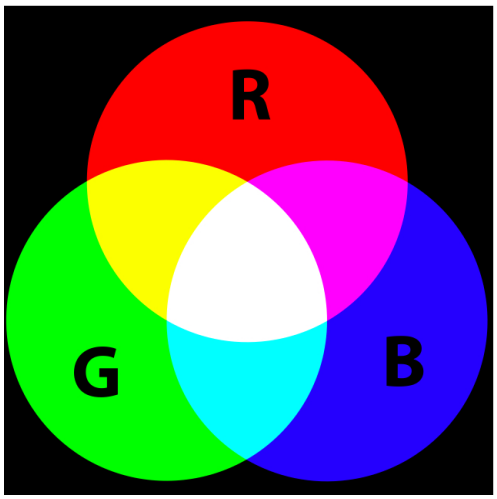

Figure 4.2: Additive Color Model

These cones are sometimes called long-wavelength cones, L-Cones, or red cones.

4.1.2 Additive Color

An additive color model involves light emitted directly from a source or illuminant. The additive process starts with black as zero, then uses a combination of red, green and blue light to form the full spectrum of visible color. Combining any two of the primary additive colors results in one of: cyan, magenta or yellow—the secondary additive colors. Combining all three yields pure white light. It should be noted that additive color is a result of the way that the eye detects color and not a property of light. There is a vast difference between yellow light, with a wavelength around 580 nm, and a mixture of red and green light; however, both stimulate our eyes in a similiar manner, so no difference is perceived.

4.1.3 Subtractive Color

If the light is reflected rather than emitted it will follow a subtractive color model. This model must be used for dyes, inks, paints, stains, etc. A subtractive color model uses the

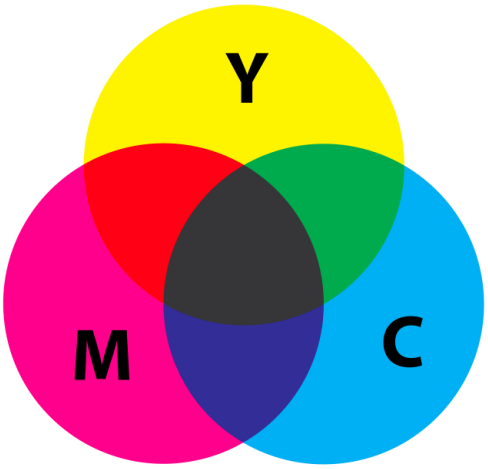

Figure 4.3: Subtractive Color Model

secondary additive colors as their basis. In this model, combining all three basic colors results in black. The zero level in a subtractive model is white and adding varying amounts of yellow, cyan and magenta yields the visible spectrum of color. This can be thought of as adding varying amounts of paint to a white canvas—the more paint, the darker the canvas becomes.

Immunohistochemical stains follow a subtractive color model, and therefore will be the model that is employed for this thesis. This is notable not only from a theoretic perspective, but also a practical one—Computers use an additive color model to represent color. The consequence of this is that all colors are given with respect to black (the assumed zero level for an additive model). To allow for the reality of subtractive color we need to change our color respresentation to use white as zero level.

4.1.4 Color Models

Since the human eye perceives color in three dimensions most mathematical models of color incorporate three dimensions. Some common computational color models are presented in the following sections.

RGB

By far the most common computational color model is RGB (*Red*,*Green*, *Blue*). This is an additive representation with (0,0,0) representing black and (1,1,1) representing white. It is represented on a computer via a three-dimensional euclidean space (a cube) as can be seen in figure 4.4. The line connecting (0,0,0) and (1,1,1) represents equal contributions of red, green and blue and thus are the grayscale values. The RGB model is convenient for computers because this is how information is displayed on monitors, televisions, projectors—all of these devices are based on an additive model of color.

RGB images are essentially captured with a monochrome camera with selective color filters. This is important for image processing because each of the channels can be represented as a grayscale image relfecting the intensities of each color. The process of combining three monochrome images to form an RGB image can be seen in figure 4.5.

RGB color is typically stored with 8-bits per channel, or collectively refered to as 24 bit color. This means that this model is capable of expressing $(2^8)^3 = 16,777,216$ different colors.

CMY

The RGB model will not work for all applications, for example, what if the user wants to print something? The printer works on a subtractive color model, not an additive one. In this case the computer would covert RGB to CMY (*Cyan*, *Magenta*, *Yellow*).

The transformation between RGB and CMY is a simple one:

$$\begin{bmatrix} c \\ m \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} r \\ g \\ b \end{bmatrix} \tag{4.1}$$

This can be seen as moving the origin from (0,0,0) black to (1,1,1) white in figure 4.4.

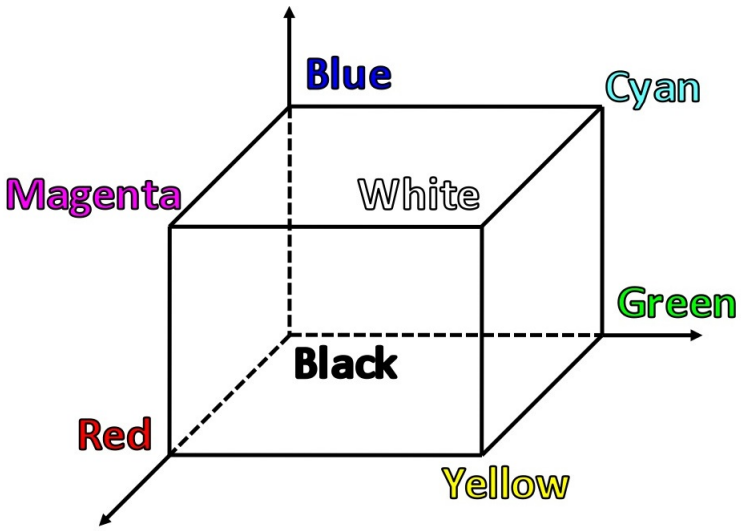

Figure 4.4: RGB Color Model

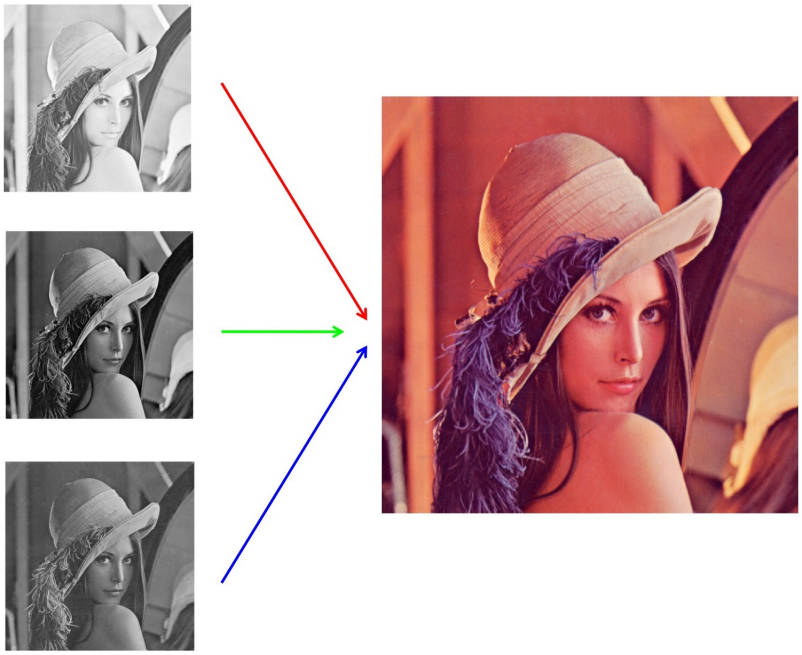

Figure 4.5: Combining 3 Channels to form a Color Image

This makes sense as dyes or inks should start at white and get closer to black as more ink/dye is added. In this case the basis vectors change from red, green and blue to cyan, magenta and yellow—the standard printing colors.

According to the above equation, equal amounts of cyan, magenta and yellow should produce black; however, in practice this isn't always the case. To account for this an additional color is added, black, giving rise to the CMYK color system. The addition of the new color dimension doesn't add any new information, but ensures high fidelity printing.

4.2 Color Clustering

Some researchers have tried different techniques to group colors in some color space in hopes of building a classifier to label each pixel as *brown* or *blue*. Conceptually, this idea is fairly simple, it groups similiar colors together—browns with browns, blues with blues, reds with reds, etc. There are static clustering techniques, methods in which static points are selected as the cluster centers and pixels are classified based on which point they are closer to. These methods work as long as there is not a great deal of variation between the target images. Unfortunately, there is a fair variation in the intensity of staining, which make static cluster centers less appealing. There are also dynamic methods used to classify pixel values. As the name suggests, these methods determine unique cluster centers for each target image. This method can deal with the variability in staining intensity but usually requires a representative sample of each color to be present in each image: if both colors are not present, unexpected results will follow. For example, if no ER positive cells are present the image will be completely stained with hematoxylin and not DAB. Since no brown is present the dynamic algorithm will not be able to find its cluster center. Color clustering techniques are generally too coarse-grained a solution for this problem, it is not enough to know that a pixel is mathematically closer to one color than another, we need to know exactly how much

each stain contributes to the color as a whole. Using the principles of a subtractive color model and the technique introduced in the following section, we can do just that!

4.3 Color Deconvolution

Color deconvolution is a technique developed to separate mixed colors which are known beforehand. It was initially developed for the problem of histological staining but was eventually adapted to other problems in forensic sciences. To the author's knowledge, this technique has not been covered by any engineering journal or textbook. It proves quite useful for certain types of color separation problems by utilizing the underlying subtractive color theory to perform a continuous separation rather than a binary one.

Two different types of color deconvolution are discussed in the literature. Ruifrok et al [16] describe a process for separating histological stains, while Berger et al [36] describe a procedure for separating ink on paper. The difference between them comes down to the difference between reflected and transmitted light. A discussion of the two methods and the law that governs the transmission of light is to follow.

4.3.1 Representation

By default, computers use an additive color representation, that is, they define black as zero and use $Red(\hat{r})$, $Green(\hat{g})$, and $Blue(\hat{b})$ as basis vectors. A typical representation plotted in 3D Euclidean space can be seen in figure 4.6.

Mathematically this can be viewed as:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = r\hat{r} + g\hat{g} + b\hat{b} \tag{4.2}$$

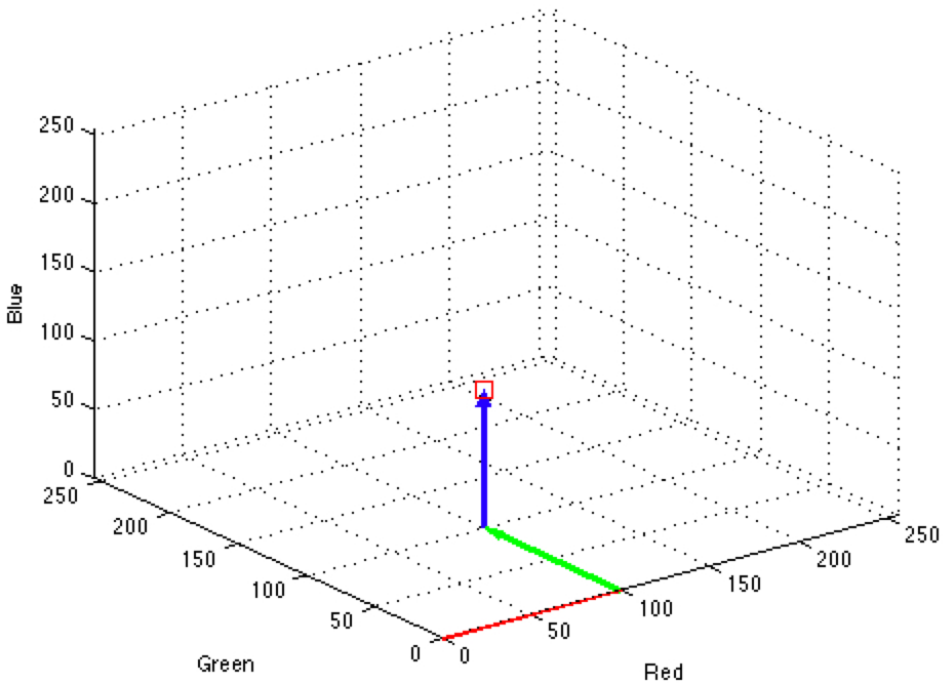

Figure 4.6: Normal Color Representation

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = r \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + g \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{4.3}$$

This is more than adequate if the colors you are interested in are red, green and blue. But what if that is not the case? This is where the concept of color deconvolution comes into play. The basic idea behind color deconvolution is changing the basis vectors from the standard $\hat{r}, \hat{g}, \hat{b}$ into colors that we are more interested in.

4.3.2 Reflected light Deconvolution

For the case of reflected light, we must first change our point of origin from black to white—as you remember, subtractive color uses white not black as its "zero" value. We make this transform by subtracting our color point from 255[1] (or whatever our background intensity is) which re-orients the color space. Once our color system has changed, we need to re-express our point in terms of colors we are interested in, for generality we shall refer to them as $\hat{c_1}, \hat{c_2}, \hat{c_3}$. Mathematically this can be expressed as:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = 255 - (c_1\hat{c_1} + c_2\hat{c_2} + c_3\hat{c_3}) \tag{4.4}$$

In this equation c_1 represents the magnitude of the unit vector $\hat{c_1}$. Herein lies the difficulty with color deconvolution: we need to define the color vectors we are interested in beforehand. We can calculate the vectors by obtaining a pure sample of one of our inks/stains and reading its standard RGB values. The point must then be re-oriented according to our subtractive color system as follows:

[1]For 8-bit color channels 255 is the maximum possible value

$$\begin{pmatrix} r_s \\ g_s \\ b_s \end{pmatrix} = \begin{pmatrix} 255 - r_a \\ 255 - g_a \\ 255 - b_a \end{pmatrix} \tag{4.5}$$

Once the point is re-oriented we need to calculate the corresponding unit vector:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \frac{1}{\sqrt{r_s \cdot r_s + g_s \cdot g_s + b_s \cdot b_s}} \begin{pmatrix} r_s \\ g_s \\ b_s \end{pmatrix} \tag{4.6}$$

We can use this approach for either two or three colors to provide the basis vectors for our equation. Below is the re-written equation in matrix form using the aforementioned basis vectors.

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = 255 - \left(c_1 \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} + c_2 \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix} + c_3 \begin{pmatrix} r_3 \\ g_3 \\ b_3 \end{pmatrix} \right) \tag{4.7}$$

$$255 - \begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \times \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{4.8}$$

The preceding equations show our original point in color space re-written in terms of our colors of interest: c_1, c_2, c_3. This can be seen graphically in figure 4.7.

Thus, if we solve for c_1, c_2, c_3, we can get the intensity of each stain (the length of the associated vector in figure 4.7).

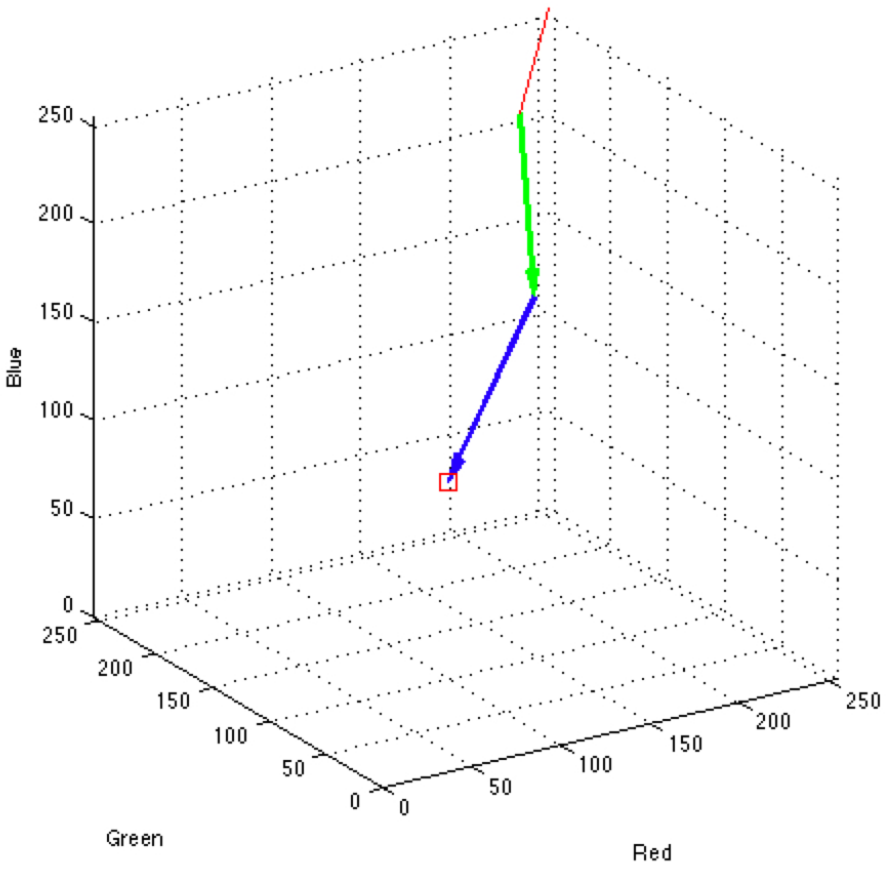

Figure 4.7: Deconvolution Color Representation

$$\begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{pmatrix}^{-1} \times \begin{pmatrix} 255 - r \\ 255 - g \\ 255 - b \end{pmatrix} \tag{4.9}$$

This method works great if you're dealing with something like ink on a page or dyes on fabric but for light microscopy we need to adjust our model to account for light travelling through the sample, rather than reflecting off of it. This is governed by the Lambert-Beer law which is discussed in the following section.

4.3.3 Lambert-Beer Law

The Lambert-Beer law states that there is a logarithmic dependence between the transmission (T), of light through a substance, the product of the absorption coefficient of the substance (α), and the distance the light travels through the material (l).

The absorption coefficient (α) can be written in terms of the product of the molar absorptivity (ϵ), and the molar concentration (c), of the absorbing material. Mathematically, this can be expressed as:

$$T = \frac{I}{I_0} = 10^{-\epsilon l c} = 10^{-\alpha l} \tag{4.10}$$

In this equation, I_0 represents the light entering the substance and I represents the light that has passed through the substance. This can be seen graphically in figure 4.8.

4.3.4 Light Microscopy Deconvolution

By applying the principles of the Lambert-Beer law to the previous discussion of color deconvolution, we can create a color model suitable for the analysis of images acquired through light microscopy. Equation 4.10 can be solved for α, which as we know is equal to $\epsilon \times l \times c$.

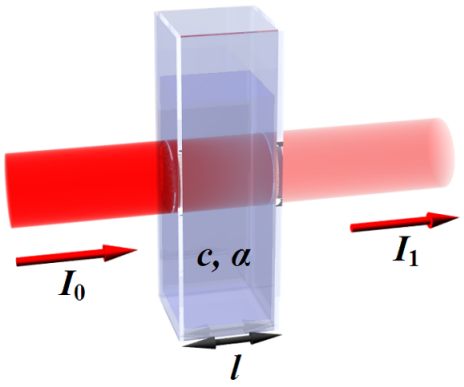

Figure 4.8: Lambert-Beer Law

Since the thickness and molar absorptivity should be constant from sample to sample, α should be linearly proportional to the concentration of stain at a given location. Solving for α yields:

$$\alpha = -\log \frac{I}{I_0} \tag{4.11}$$

In our case we assume the light entering the substance (I_0), is equal to 255. As we know from the previous sections this is the value assigned to white light in an additive color model.

Similiar to the light reflection case we must know the RGB values of our colors of interest, and calculate them in an analogous fashion. First, we pick a point representing a pure stain and calculate its light transmissivity:

$$\begin{pmatrix} r_t \\ g_t \\ b_t \end{pmatrix} = \begin{pmatrix} \log(\frac{r_a}{255}) \\ \log(\frac{g_a}{255}) \\ \log(\frac{b_a}{255}) \end{pmatrix} \tag{4.12}$$

Once the point is transformed we need to calculate the corresponding unit vector:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \frac{1}{\sqrt{r_t \cdot r_t + g_t \cdot g_t + b_t \cdot b_t}} \begin{pmatrix} r_t \\ g_t \\ b_t \end{pmatrix} \tag{4.13}$$

Once we have calculated our unit vectors of interest, we can proceed to re-express colors in the image in terms of our basis colors. In the following equation, r, g, b indiciates the point we wish to transform, while r_1, r_2, r_3 correspond to the unit vectors we calculated using equation 4.13.

$$\begin{pmatrix} \log(\frac{r}{255}) \\ \log(\frac{g}{255}) \\ \log(\frac{b}{255}) \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{pmatrix} \times \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{4.14}$$

Solving for c_1, c_2, c_3 yields:

$$\begin{pmatrix} c_1 \\ c_2 \\ c_2 \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{pmatrix}^{-1} \times \begin{pmatrix} \log(\frac{r}{255}) \\ \log(\frac{g}{255}) \\ \log(\frac{b}{255}) \end{pmatrix} \tag{4.15}$$

Using this equation it becomes possible to quantify the level of staining at any given point in a digital histological image. As noted previously, when written in terms of transmissivity, the stains add linearly and increase at a linear rate. For example, figure 4.9 shows linearly spaced points along the unit vector defined for our brown stain (DAB). If we transform these points back into normal RGB space as seen in figure 4.10 we can see that the linear relationship is lost. This is an important point as it shows the limitation of analyzing stain values in RGB space—not only do the stains not form nice clusters, they do not even follow a linear relationship.

Figure 4.9: DAB Stain Spectrum

4.3.5 Calculating the third Color

The astute reader may have noticed that all the preceeding equations require three color vectors while our color problem has only two colors—so where does the third one come from? Both Ruifrok et al [16] and Berger et al [36] describe a procedure for calculating the third vector. Ruifrok et al [16] uses the following equation to calculate the third vector:

$$r_3 = \sqrt{1 - r_2^2 - r_1^2} \tag{4.16}$$

The idea behind this is that the third color should be whatever is left of the RGB spectrum. So, simply speaking, if the first and second colors were mostly blue and green respectively the third color would be mostly red. While this sounds reasonable it does not work terribly well in practice. The problem seems to be that the third color is not really a color but an error term. Theoretically, every pixel in the image should expressable as a

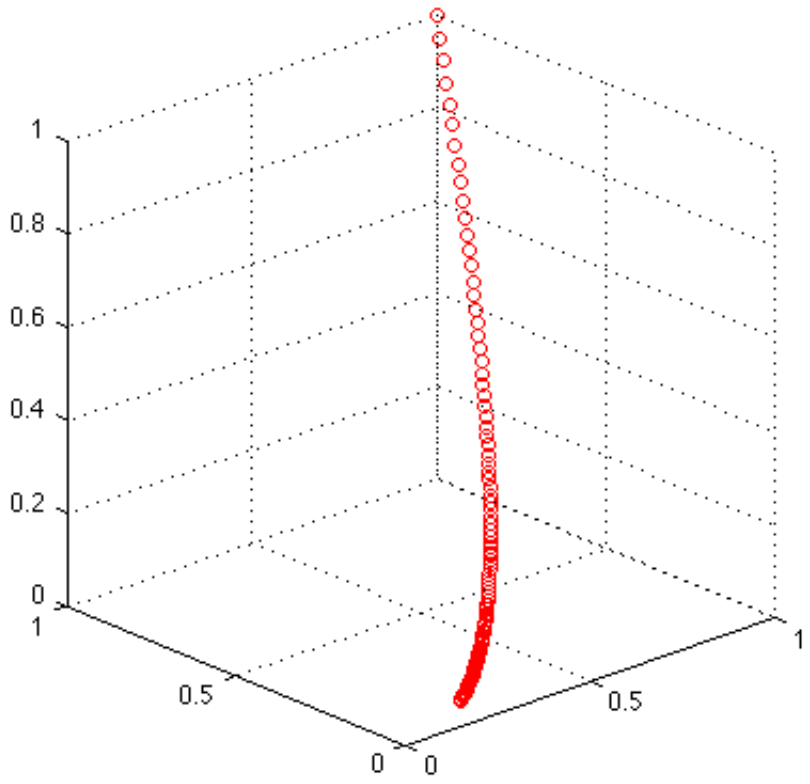

Figure 4.10: Stain Intensities plotted in RGB

combination of the two stain color vectors (blue and brown in this case). In reality the image is never this perfect, there is always artifacts within the images, variable lighting conditions etc. Visually, all the pixels should lie on the plane defined by the two color vectors, keeping this in mind it would be reasonable to define error as the distance the pixel in question lies from our ideal plane. Mathematically this can be accomplished with a cross product.

$$\begin{pmatrix} r_3 \\ g_3 \\ b_3 \end{pmatrix} = \begin{pmatrix} r_1 \\ g_1 \\ b_1 \end{pmatrix} \times \begin{pmatrix} r_2 \\ g_2 \\ b_2 \end{pmatrix} \tag{4.17}$$

This was the approach taken by Berger et al [36] and makes more sense for analyzing histological sections with only two stains.

Chapter 5

Classification

This chapter will describe the process employed to determine if a cell is, in fact, estrogen receptor postive or negative. All cells should be stained by hematoxylin, giving them a blue-ish appearance. The addition of DAB should stain estrogen receptor positive cells, giving them a brown hue. The chapter on color separation provided a discussion of how these colors can be separately quantified. Once separated we can use the quantified hematoxylin and DAB staining to classify a cell as either ER negative or ER positive.

We use the segmentation discussed in chapter 3 to isolate the nuclei of the epithelial cells. Once the nuclear regions were isolated the color data was extracted using the previously described color deconvolution algorithm. We then calculate the average brown stain and blue stain for the nuclear pixels and plot that as a point in two dimensional space. This is done for each of the segmented nucleus and we are left with a plot like the one seen in figure 5.1. The blue line represents an example of a linear discriminant, that is, all cells above the line are predominantly brown (ER positive) and all points below the line are predominantly blue (ER negative).

Questions that will be addressed in this chapter revolve around the aforementioned graph. Firstly, how do we determine the optimal linear discriminant? After we have determined

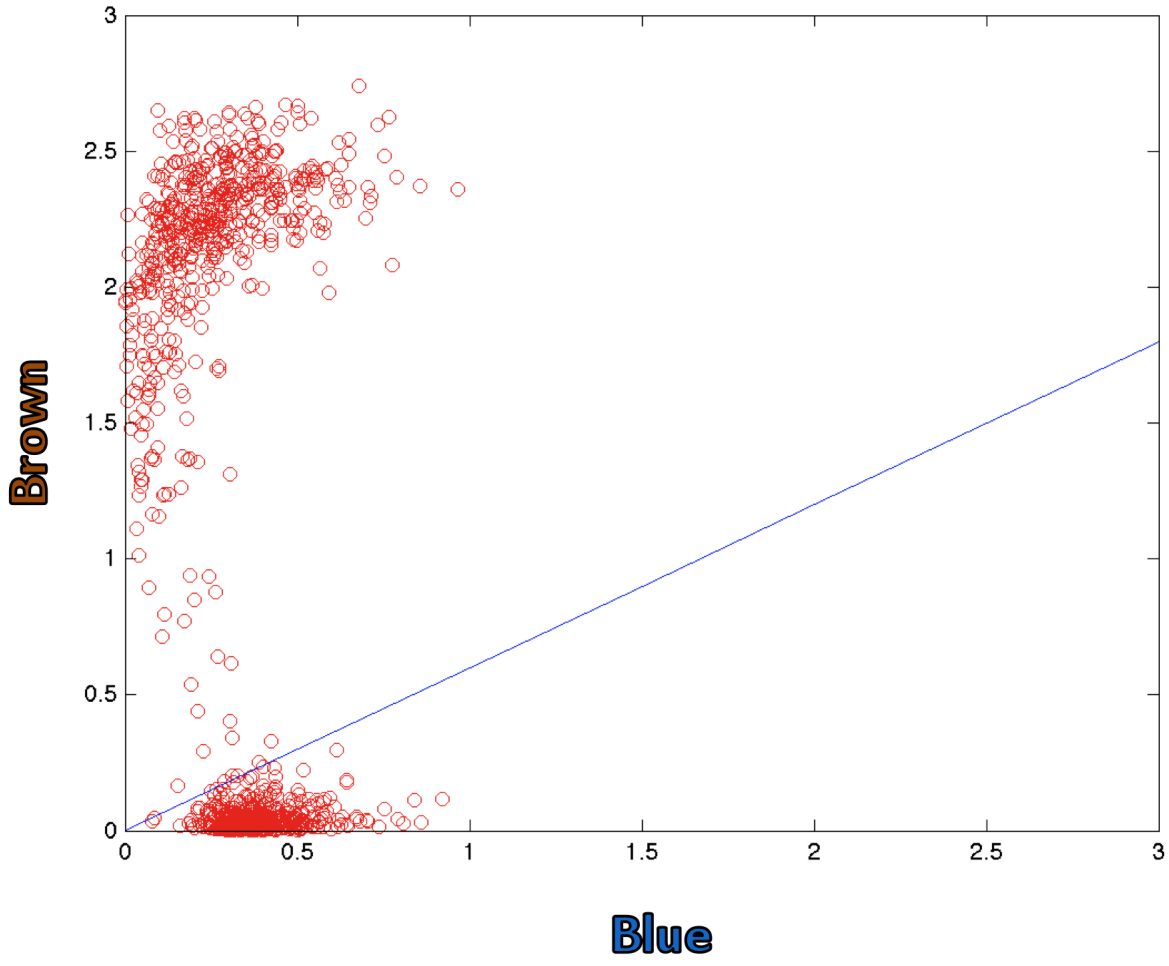

Figure 5.1: Sample Cell Color Plot

that, what percentage of cells are required to be ER positive before the tumor itself is deemed to be ER positive. To answer this question we will need to review some basic biostatistics and introduce the receiver-operator curve.

5.1 Linear Discriminants

A linear discriminant function is a linear combination of $\mathbf{x}$ [37]. In our case $\mathbf{x}$ is a vector composed of the blue and brown staining intensities, or mathematically:

$$\mathbf{x} = \begin{pmatrix} br \\ \\ bl \end{pmatrix} \tag{5.1}$$

The general form of the linear discriminant of $\mathbf{x}$ can be written as:

$$g(x) = \mathbf{w^t x} + w_0 \tag{5.2}$$

Where $\mathbf{w^t}$ is the vector of weights specifying the location of the linear discriminant. A two-category case follows the following rule: if $g(\mathbf{x}) > 0$ then $\mathbf{x}$ belongs to Class 1 (Estrogen Receptor Positive) or if $g(\mathbf{x}) < 0$ then $\mathbf{x}$ belongs to Class 2 (Estrogen Receptor Negative). Graphically, this corresponds to above and below the classification line. So for the graph seen in figure 5.1, equation 5.2 defines the blue line and all the cells above the line are estrogen receptor positive and all those below are estrogen receptor negative.

Two major cases exist when dealing with linear discriminant functions: is the data linearly separable? Suppose we have two sample sets in two-dimensional space, if a line exists that can classify all of the samples correctly the samples are said to be linearly separable. More generally, two points are said to be linearly separable in n-dimensional space if they can be separated by a hyperplane. In practice, one would only use an algorithm for linearly separable data if there was a reason to believe that the error rate for the optimal linear discriminant would be low. Unfortunately, large data sets are almost always non-linearly separable. For this reason we will assume our data to be non-linearly separable and use an appropriate algorithm.

Training a classifier requires labelling the data ahead of time as either ER positive or ER

negative and then using an alogrithm to determine the best linear discriminant to classify future cases. The cells were visually assessed by a trained pathologist as predominantly brown or blue and then labelled as such. This labelled data along with the algorithm described in the following section were used to calculate an appropriate linear discriminant function.

5.1.1 Minimum Squared Error Approach

Assuming we have a set of n d-dimensional (d=2 in our case) samples $\mathbf{x_1}, ..., \mathbf{x_n}$, n_1 of which are in subset D_1 (estrogen receptor positive) and n_2 are in subset D_2 (estrogen receptor negative). We define the variable y_i as the x_i values along with a threshold component, $x_0 = 1$ to make an augmented pattern vector. Further, if the sample is labeled as D_2 (ER negative), then the entire pattern is multiplied by -1. More formally this can be expressed as:

$$Y = \begin{pmatrix} 1 & X_1 \\ -1 & -X_2 \end{pmatrix} \tag{5.3}$$

Where X_1 represents the color vector for an estrogen receptor positive cell and X_2 represents the color vector for an estrogen receptor negative cell. If Y were square and non-singular we could write:

$$a = Y^{-1}b \tag{5.4}$$

Using this form we could obtain a formal solution immediately. However, in most practical cases Y will be rectangular, usually with far more rows than columns ($n >> d$). Therefore, an exact formal solution is impossible. We can, however, seek a weight vector $\mathbf{a}$ that minimizes the error between $\mathbf{Ya}$ and b. We can define the error vector $\mathbf{e}$ by:

$$e = Ya - b \tag{5.5}$$

One approach to minimizing $Ya - b$ is to try to minimize the square length of the error vector (e), that is, the minimum squared error. It can be shown [37] that minimizing the square error requires that:

$$\mathbf{Y^t Ya = Y^t b} \tag{5.6}$$

Solving for $\mathbf{a}$ yields:

$$\mathbf{a} = (\mathbf{Y^t Y})^{-1}\mathbf{Y^t b} \tag{5.7}$$

$$\mathbf{a} = \mathbf{Y}^{+}\mathbf{b} \tag{5.8}$$

Where $\mathbf{Y}^{+}$ is defined as $(\mathbf{Y^t Y})^{-1}\mathbf{Y^t}$ and is called the *pseudoinverse* of $\mathbf{Y}$. Using this approach we can always find a solution which minimizes the classification error. It is worth noting that if the data is truly linearly separable this algorithm does not guarantee finding such a solution.

5.1.2 Implementation

The aforementioned algorithm was developed using MATLAB along with a simple program to label segmented nuclei as either ER positive or negative to provide the data to build the classifier. The cells were labelled as positive or negative by visual inspection performed by a pathologist. A good cross section of cells was used (strongly positive and negative as well as borderline) to optimize the classifier. Over 500 cells were used to train the classifier, with approximately 10% being strongly positive, 10% being weakly positive, 10% weakly negative

and the remainder as strongly negative.

5.2 Basic Biostatistics

This section is intended to familiarize the reader with some basic concepts in biostatistics. Most of these concepts exist in other fields, albeit under different names. All of the following are statistical performance measurements for binary classifications. For a more complete discussion of biostatistics and research methodologies see Keppel et al [38].

5.2.1 Sensitivity

Sensitivity, also known as true positive rate, relates a test's ability to identify positive results. It answers the question, "What is the probablity this person will have a positive test given they have condition X". More formally this can be expressed as $P(PositiveTest|ConditionX)$. Sensitivity can be calculated using the following equation:

$$Sensitivity = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{5.9}$$

5.2.2 Specificity

Specificity, also known as true negative rate, relates a tests ability to identify negative results. It answers the question, "What is the probability this person will have a negative test given they don't have condition X". More formally this can be expressed as $P(NegativeTest|\neg ConditionX)$. Specificity can be calculated using the following equation:

$$Specificity = \frac{TrueNegatives}{TrueNegatives + FalseNegatives} \tag{5.10}$$

5.2.3 Positive Predictive Value

Positive predictive value relates the proportion of positive tests that will actually be true positives. It answers the question, "What is the probability this person with a positive test actually has condition X". More formally this can be expressed as $P(ConditionX|PositiveTest)$. Positive predictive value can be calculated using the following equation:

$$PPV = \frac{TruePositives}{TruePositives + FalsePositives} \tag{5.11}$$

5.2.4 Negative Predictive Value

Negative preditive value relates the proportion of negative tests that will actually be true negatives. It answers the question, "What is the probability this person with a negative test doesn't have condition X". More formally this can be expressed as $P(\neg ConditionX|NegativeTest)$. Negative predictive value can be calculated using the folowing equation:

$$NPV = \frac{TrueNegatives}{TrueNegatives + FalseNegatives} \tag{5.12}$$

5.3 Receiver-Operator Curves

A receiver operator curve, or simply ROC curve, is a graphical plot of sensitivity (true positive rate) vs. 1-specificity (true negative rate) for a binary classification system as its classification threshold is varied. In this particular case that would mean varying the percentage of cells required for classification as ER positive from `0%` to `100%`. For each increment the sensitivity and specificity is re-calculated and plotted on the ROC curve.

A sample ROC curve is seen in figure 5.2. The blue line represents a sample ROC curve from a reasonably good test—i.e. high true positive rate, low true negative rate. The red line represents a ROC curve that would be generated from a test based purely on chance—i.e.

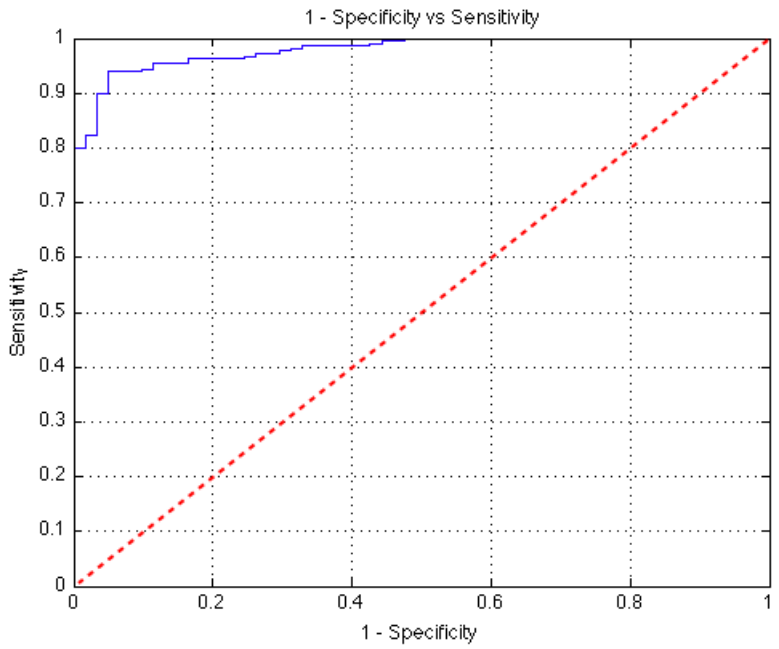

Figure 5.2: Sample ROC curve

flipping a coin to determine ER status. The area under a ROC curve has been shown to be a good measure of a test's validity.

ROC curves were first employed by electrical engineering in the 1940's and 50's for finding objects in radar. Since then these curves have become extensively used in clinical medicine in selecting decision thresholds and comparing the efficacy of different tests to classify the same condition/outcome.

5.3.1 Optimal Threshold from ROC

It is possible to use the ROC curve to find the optimal cut point, in this case the percentage of estrogen receptor positive cells needed to deem the cancer estrogen receptor positive. This approach calculates the point on the ROC curve closest to (0,1). This will simultaneously maximize both the sensivity and specificity of the test in question. The distance to (0,1) can be calculated using each sensitivity, s_n, and specificity, s_p as follows:

$$d = \sqrt[2]{(1 - s_n)^2 + (1 - s_p)^2} \tag{5.13}$$

The sensitivity and specificity corresponding to the minimum distance are the optimal values. The optimal cut point is that which yields the optimal sensitivity and specificity. It should be noted that this approach weights sensitivity and specificity equally, and this isn't always appropriate. For example, in some cases the cost associated with a false negative may be much higher than a false positive. This would require a more sophisticated approach and will be addressed further in the dicussion section.

Chapter 6

Results and Discussion

The preceding chapters have discussed the individual components of the image analysis systems and their respective mechanisms. This chapter will outline the results obtained using the two separate data sets.

The chapter will conclude with a discussion of the future directions of this research.

6.1 Results

6.1.1 Data Set I - MUN

The slides acquired from Eastern Health were part of the internal quality assurance study done by Dr. Makretsov (assistant professor, Faculty of Medicine, Discipline of Pathology) in 2008-2009 for Eastern Health. The slides were not prepared for diagnostic purposes; tumor tissue blocks were used after the diagnostic tests were completed. This was done for internal quality assurance purposes in order to make a comparison with external laboratory results only (to make sure that internal tests are adequate and can be offered to the patients in the future). These slides were not part of patient material or patient records as the clinical ER test was done and reported by an external laboratory (and constitute clinical

	Sensitivity	Specificity	PPV	NPV
Morphology Algorithm	100	100	100	100

Table 6.1: Results of Eastern Health Images

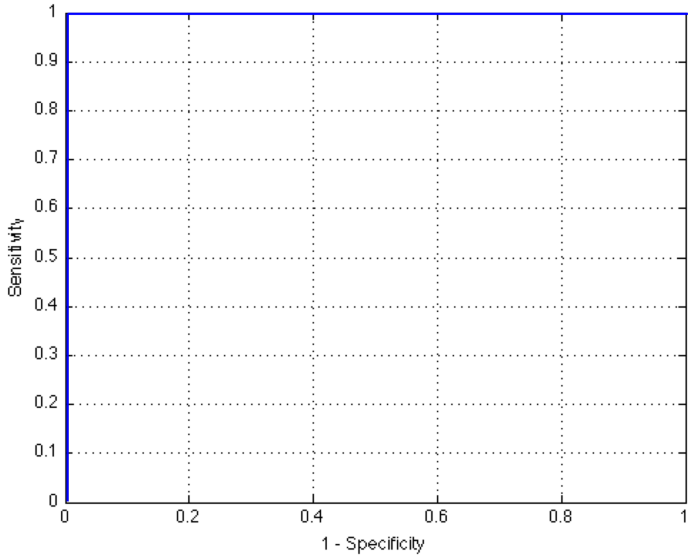

Figure 6.1: ROC curve for MUN Images

results and clinical records). Ergo our slides were considered internal QA (quality assurance) slides only. The slides were scanned using Aperio's histological slide scanner. The images were scanned into a proprietary format and were too large to process on their own (2GB per image). Representative sections containing malignant cells had to be captured using a screen capture tool. These captures were then reviewed by a trained breast pathologist to ensure they contained a sufficient number of malignant cells. Once the images were verified they were then run using the algorithms discussed in previous chapters. A receiver operator curve (ROC) was generated to determine the optimal cut point, which was 3% for the Eastern Health images. The classification data can be seen in Table 6.1 and resulting ROC curve can be seen in figure 6.1.

6.1.2 Data Set II - UK Data

The 540 breast carcinomas were taken from a population based case series from East Anglia, UK. A 0.6mm core was taken from each formalin-fixed, paraffin embedded donor block and used to construct a tissue microarray. The Ariol (Applied Imaging Inc.) slide scanner was used for digitalization of the stained slides and all slides were scanned in high resolution at 200x optical magnification.

All slides were reviewed and scored by a pathologist specializing in breast cancer. The images were analyzed by the morphology-based algorithm as described in the preceding chapters, as well as ARIOL and ImageJ systems (as described below).

A receiver operator curve was generated to determine the optimal cut point, which was `6.5%` for the UK images. The classification results for the morphology-based system, ARIOL and ImageJ are listed in Table 6.2. The resulting ROC curve can be seen in figure 6.2.

ARIOL

ARIOL is a system developed by Leica Microsystems that includes both a digital slide scanner and quantitative image analysis software for immunohistochemical stained slides.

ImageJ

ImageJ is a freely available java-based image processing tool developed by the National Insititute of Health. There are numerous plug-ins available to provide a wide range of functionality. The analysis employed a simple thresholding by Otsu's[33] method followed by Ruifrok's[16] color deconvolution algorithm.

	Sensitivity	Specificity	PPV	NPV
Morphology Algorithm	94	95	99	82
Ariol	85	97	99	64
ImageJ	80	84	95	54

Table 6.2: Results of Cambridge Images

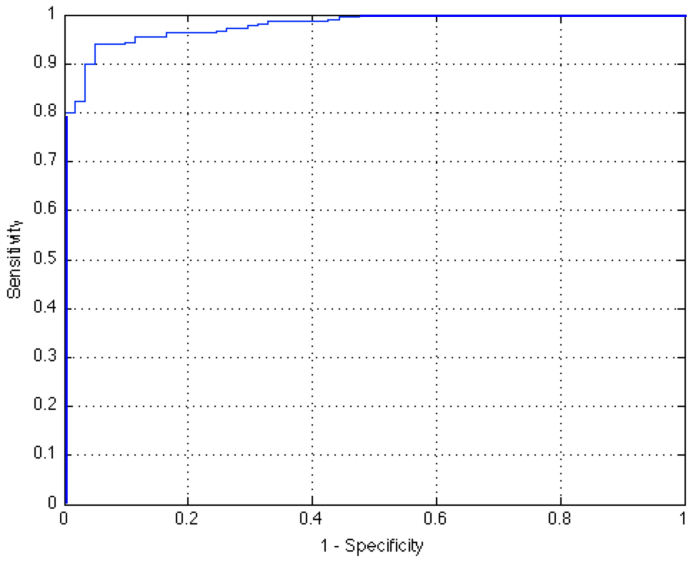

Figure 6.2: ROC curve for Cambridge Images

6.2 Discussion

The results obtained in this study are promising; they demonstrate that morphology-based algorithms are able to achieve superior classification results compared with exisiting commercial software and freely available research-inspired software. The most striking difference would be that of sensitivity: there was almost a 10% improvement in sensitivity over the ARIOL and imageJ systems. This means that this system is better able to identify the truly estrogen receptor positive biopsies.

This highly sensitive system could be of benefit after the recent Cameron Inquiry: an investigation into the erroneous estrogen receptor tests done between 1997 and 2005. Retesting of biopsies taken during this time revealed 383 patients who had received incorrect test results. A large portion of these patients were falsely deemed estrogen receptor negative and were therefore denied the adjuntive hormone therapy *tamoxifen*. The proposed system could be used as an internal quality control measure to verify the pathologist's report. If there was any discrepancy between the pathologist and the automated system it could be reviewed by a second, independent pathologist. Furthermore, components of the system could be used for other quality control measures—for example, using the color deconvolution algorithm to measure average staining intensities, to ensure a consistent staining level between samples.

The falsely classified biopsies are the result of inadequate segmentation. If, for example, the algorithm fails to segment a clump of positive cells it can dramatically skew the count and ultimately the final classification. This is a potential area for improvement and will be further discussed in future directions.

Finally, although there are some fully-automated aspects to the system, it does require calibration of cell size and color vectors for each data set. Pre-set values could be set based on industry averages, but unfortunately there is significant variation between laboratories. Further work on auto-calibration is dicussed in future directions.

6.3 Future Direction

In future work an effort should be made to develop a more fully-automated system. Currently there are several areas in which the user must intervene to calibrate the system for a new set of images. The work of Rabinovich et al[39], in particular using non-negative matrix factorization[40], showed a promising method to automatically determine the basis color vectors used in the staining process. As discussed in the color separation chapter, these vectors are required for the color deconvolution and the user must determine them manually. An algorithm was developed for this system using the non-negative matrix factorization approach along with the color deconvolution algorithm described in Ruifrok et al[16] with sub-optimal results—that is, results inferior to the manual approach. It is possible that if the deconvultion algorithm described in section 4.3.5 were used with non-negative matrix factorization more appropriate vectors would be generated. Similiarly, the choice of structuring elements—see Appendix A—could be automated by using techniques described by Soille[32]. He advocates using a series of incrementally larger structuring elements to determine the average object size.

Finally, the segmentation algorithm could be improved upon in future versions. While the current algorithm could segment the majority of cells correctly it would often miss some, especially clusters of cells that the algorithm was unable to split. This problem was of little consequence in most of the samples, but was responsible for misclassification of samples when a cluster of positive cells were missed. A more robust segmentation algorithm could therefore improve the overall performance of the system.

Ultimately the results are only as good as the pathologist's scoring, which we are using as the ground truth data.To truly assess how the system performs against a trained breast pathologist one would need the clinical outcomes associated with each tissue sample. That is, did the patient survive the cancer? If not, how long did they live before they sucumbed

to their disease? Using this data and a Kaplein-Meyer analysis we could see whether the patholgist or the software is a better predictor of long term outcomes. If further work were to be done on this system it would be tremendously helpful to analyze the clincial outcomes to get an absolute sense how the software compares with the pathologists' ratings.

Appendix A

Mathematical Morphology

Mathematical morphology is a theory and technique for analysis of spatial structures based on set theory, lattice theory and integral geometry. It is most often applied to digital images, but is not limited to them.

Mathematical morpholgy was born of the study of porous media in the mid-sixties in France. Porous media are binary in the sense that a point either belongs to a pore or the matrix surrounding the pore. This led G. Matheron and J. Serra to introduce a set formalism for analyzing binary images. If the image was to be considered a set of object points and a set of background points the image could be processed using simple operations such as unions, intersections, complementation and translations.

During the 1960's and 1970's mathematical morphology dealt with primarily binary images and generated a large number of binary operators: erosion, dilation, opening, closing, thinning, skeletonization, ultimate erosion, etc.

During the late 1970's and 1980's the theory was extended to include grayscale images. By viewing grayscale images as topological surfaces, researchers were able to extend the notion of the basic operators (erosion, dilation, etc), but also introduce new operators such as morphological gradients, top-hat transforms and the watershed transform (powerful seg-

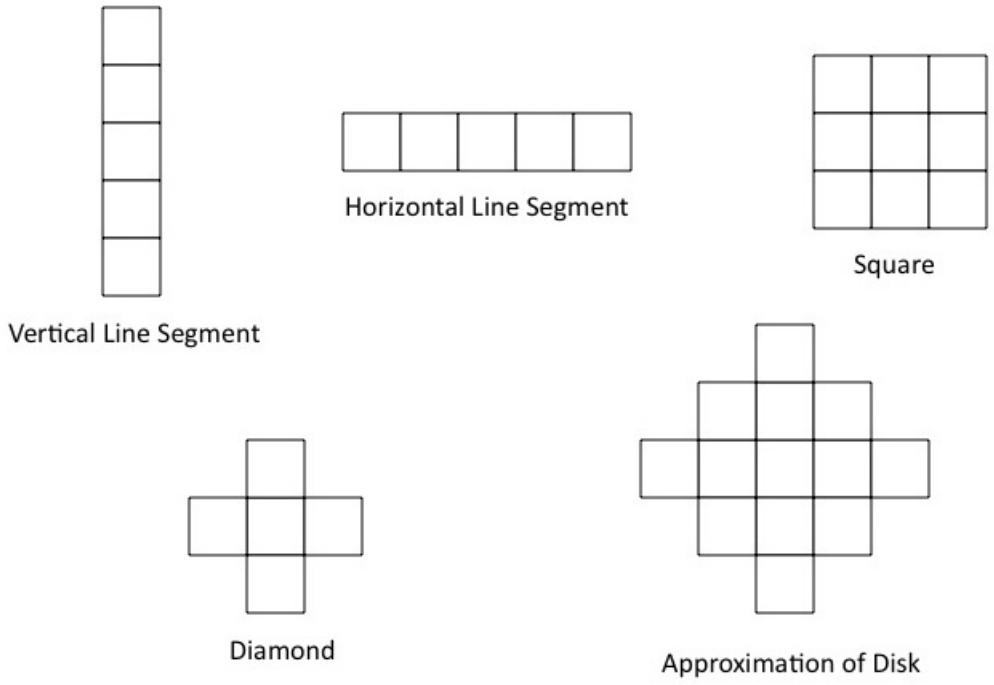

Figure A.1: Sample Structuring Elements

mentation algorithm to be discussed later).

Over the past two decades, mathematical morphology has seen more advancements in theory and a large increase in popularity—this is probably due to the advent of more efficient algorithms [34] allowing a much wider range of possible applications.

A.1 Basic Operations

Morphological operators aim at extracting important structures from an image. This is achieved by probing the image with another set of known shape called the *structuring element*. The structuring element is chosen based on a priori knowledge of the desired structure and the undesired structures—can be noise or just objects in the image that we want to suppress.

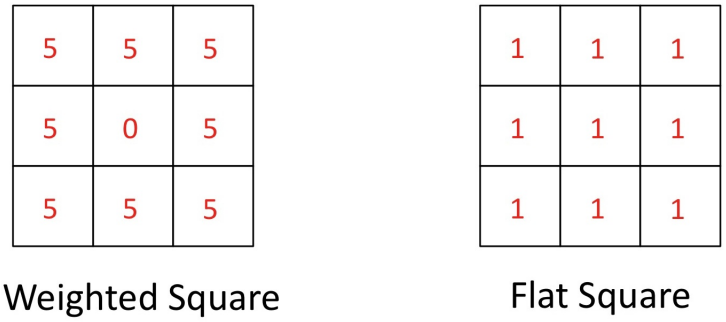

Figure A.2: Weighted and Flat Structuring Elements

Structuring elements can be round, straight lines, squares, diamonds, or any other shape as can be seen in figure A.1. A vertical line structuring element could be used to enhance vertical lines in an image while suppressing other structures. The size of the structuring element can be changed depending on the size of the desired object.

Addtionally, structuring elements can have weights or heights associated with each pixel. A structuring element with equal weights is referred to as a *flat* structuring element. An example of a flat and weighted structuring element can be seen in figure A.2. Weighted structuring elements are beyond the scope of this thesis, but can be used in situations in which more complex structures need to be indentified. For example, if we were looking at a cytoplasmic stain rather than a nuclear stain a weighted structuring element could help identify the ring of stain surrounding a cell. For a more complete discussion of weighted structuring elements please see Soille [32].

A.2 Erosion

In its simplest terms, an erosion operation on a binary image can be thought of as *Does the structuring element fit the set?*.

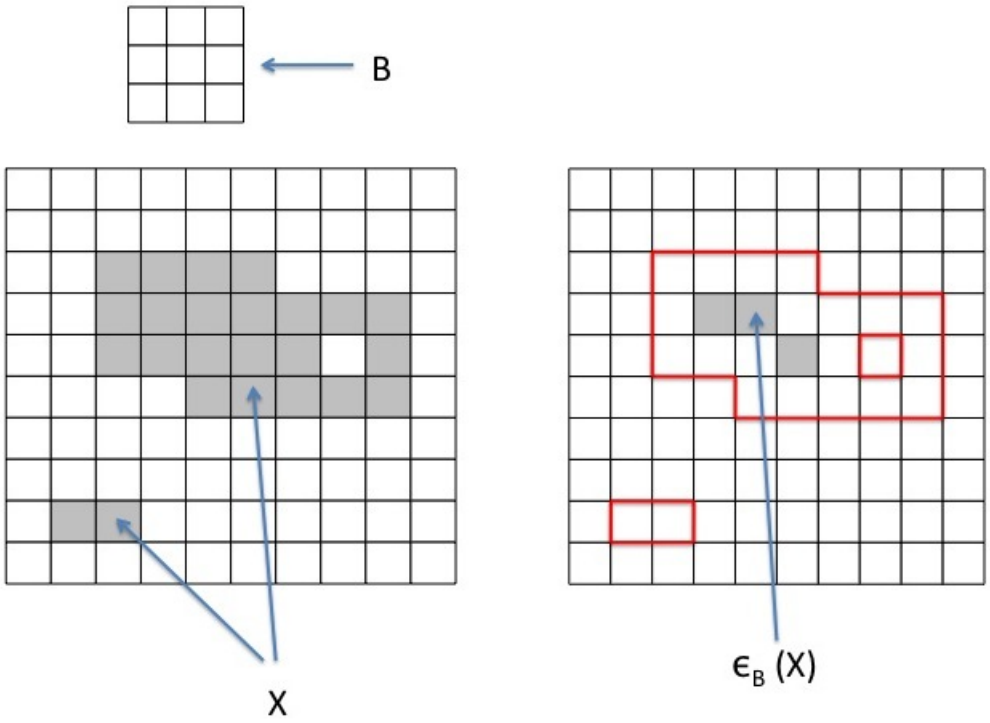

Figure A.3: Binary Erosion Example

The erosion of a set X by a structuring element B is denoted by $\epsilon_B(X)$ and is defined as the set of points x such that B is included in X when its origin is placed at x:

$$\epsilon_B(X) = \{x | B_x \subseteq X\} \tag{A.1}$$

An example of this can be seen in figure A.3. This idea can be extended to grayscale images as follows:

$$[\epsilon_B(f)](x) = \min_{b \in B} f(x + b) \tag{A.2}$$

This equation basically means that the pixel at x will be set to the lowest pixel value in the neighbourhood, as defined by the size and shape of the structuring element B.

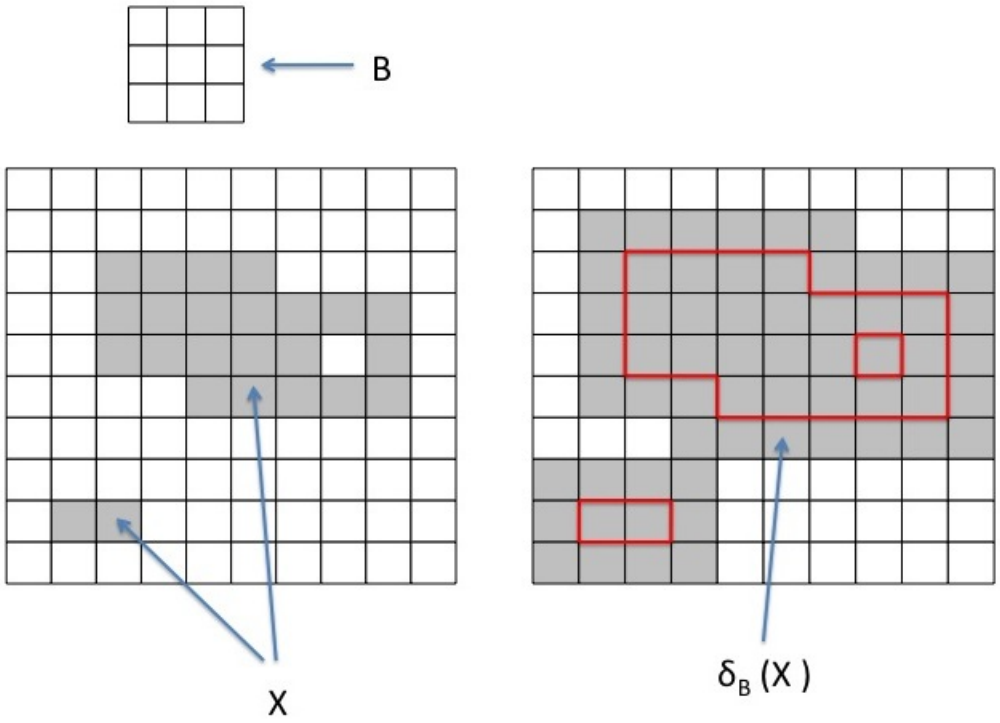

Figure A.4: Binary Dilation Example

A.3 Dilation

The dilation operator can be thought of as the opposite of the erosion operator. In its simplest terms, the dilation operation on a binary image can be thought of as *Does the structuring element hit the set?*.

The dilation of a set X by a structuring element B is denoted $\delta_B(X)$ and is defined as the set of points x such that B hits X when its origin coincides with x:

$$\delta_B(X) = \{x | B_X \cap X \neq \emptyset\} \tag{A.3}$$

An example of this can be seen in figure A.4. This idea can be extended to grayscale images as follows:

$$[\delta_B(f)](x) = \max_{b \in B} f(x + b) \tag{A.4}$$

This equation basically means that the pixel at x will be set to the highest pixel value in the neighbourhood, as defined by the size and shape of the structuring element B.

A.4 Opening

The erosion of an image not only removes objects smaller than the structuring element, but also shrinks all other objects. This problem is solved by dilating the image previously eroded using the same structuring element. This is the definition of the *opening* operation—or more formally:

$$\gamma_(f) = \delta_B[\epsilon_B(f)] \tag{A.5}$$

All the structures will not be recovered; those removed completely by the erosion operation will not be recovered. Herein lies the power of the opening operator: image structures are selectively filtered out based on the size and shape of the structuring element.

An example of a closing operation on a binary image can be seen in figure A.5.

Just as with the erosion and dilation operators, using the above definitions, the opening operation can be applied to grayscale images. This will remove small isolated areas of high intensity—this could be noise or even cell nuclei.

A.5 Closing

The *closing* operator is the dual of the opening operator. It is intended to recover the shapes of objects that have been dilated. Unlike the opening operator which filters objects, the closing operator filters holes within objects or spaces between objects. So for example,

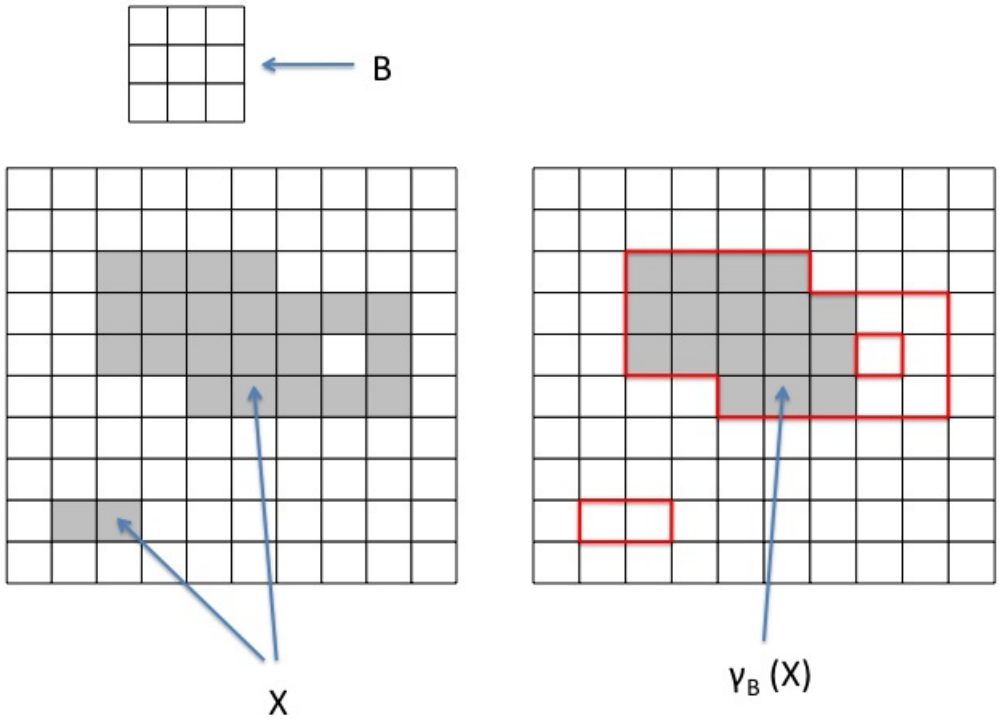

Figure A.5: Binary Opening Example

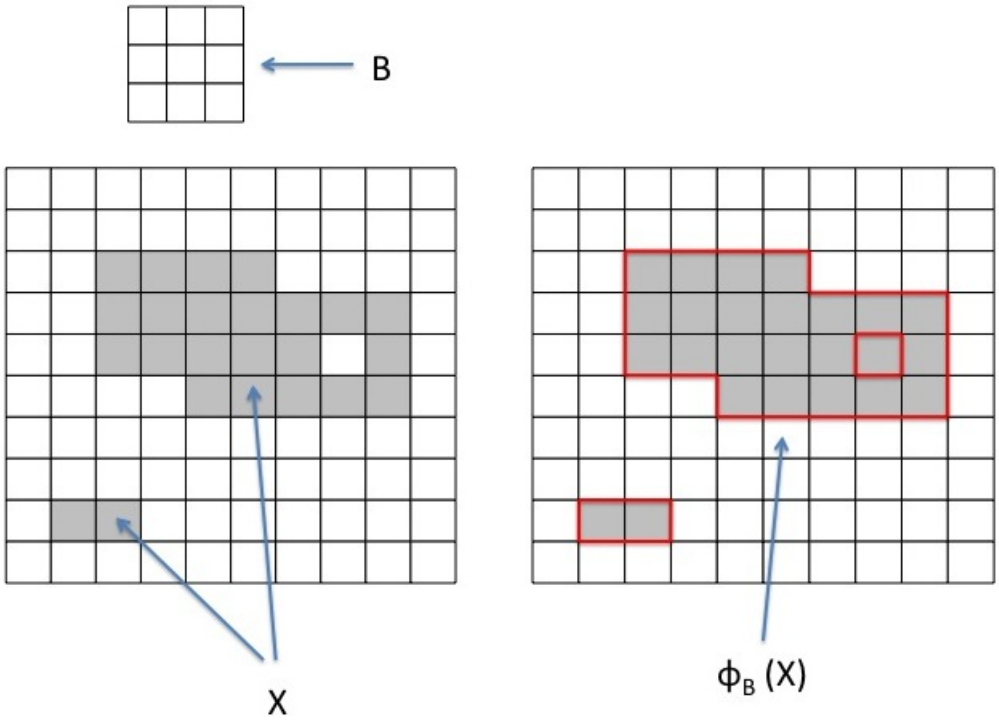

Figure A.6: Binary Closing Example

if you had a binary object with a hole smaller than the structuring element it will be erased by the dilation and the subsequent erosion will not be able to recover it. More formally the closing operation is defined as:

$$\phi_B(f) = \epsilon_B[\delta_B(f)] \tag{A.6}$$

An example of a closing operation on a binary image can be seen in figure A.6.

As with the opening operation the closing operation can be used for grayscale images if the grayscale defintions of erosion and dilation are used. This will remove isolated areas of low intensity that are contained within a region of high intensity.

A.6 Top Hat

Top hat operations are based on the idea that it is often easier to remove structures of interest than to suppress all other structures. Structures can be selectively filtered out based on the the size and shape of the structuring element, but what if these are the structures of interest? These structures can be recovered by taking the arithmetic difference between the original image and its opening or closing.

A white top-hat (or top hat by opening) is defined as the difference between the original image f and its opening γ, or more formally:

$$WTH(f) = f - \gamma(f) \tag{A.7}$$

Similiarly, the black top-hat (or top hat by closing) is the difference between the closing of the original image and the original image:

$$BTH(f) = \phi(f) - f \tag{A.8}$$

Both of these transformations can be applied to grayscale images using the grayscale definitions of opening and closing. The principle is the same, isolated areas of high or low intensity (opening vs. closing)

A.7 Reconstruction

Reconstruction operations are based on *geodesic transformations*, they differ from all the previously discussed morphological operators in that they require two input images rather than one. A morphological operation is applied to one image while being forced to remain above or below the second image.

An example of a geodesic dilation can be seen in figure A.7.

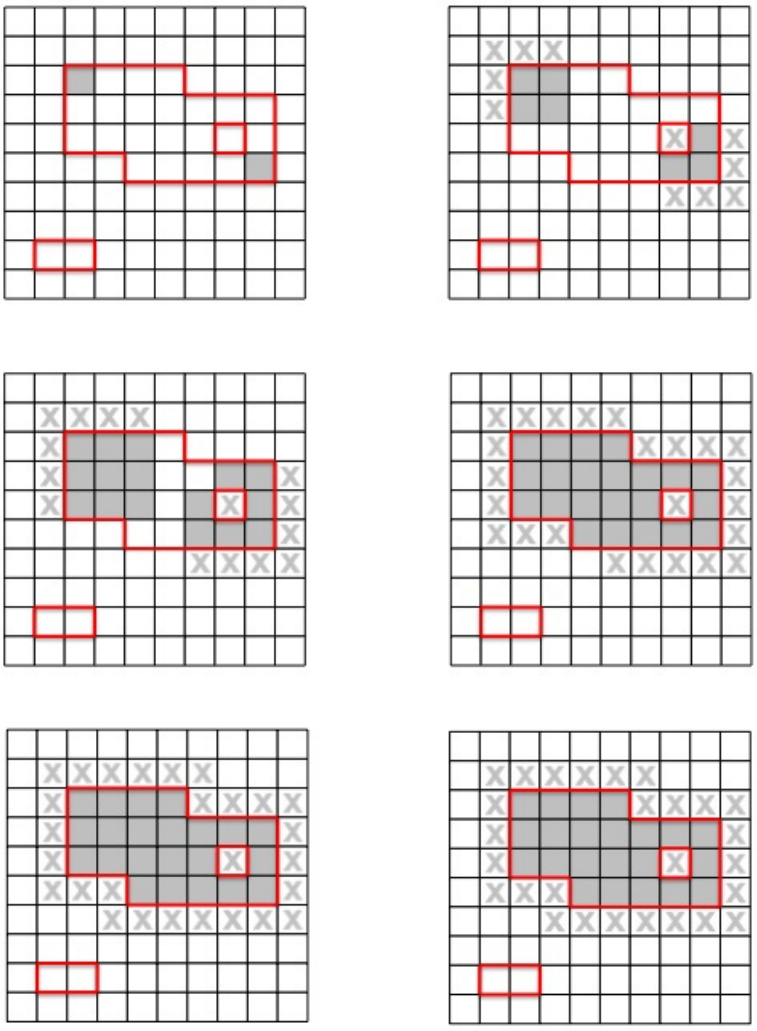

Figure A.7: Binary Geodesic Dilation Example

In the first image we have one image overlaid on top of the other—the first being the gray pixels and the second being the red outline. The structuring element for this example is the same as previous examples. The gray X's are included to show where the first image was dilated but couldn't be included in the final image because of being constrained to within the red lines. Shown in this example are successive dilations; if the operation is continually applied it will eventually reach an equilibrium—that is, no further geodesic dilations will change the image.

Geodesic operations are rarely used on their own, but when they are iterated to stability they lead to a class of operations known as *Reconstruction Operations*. There exist reconstruction-type versions of all of the above mentioned morphological operators, they will not be discussed here. For a more complete discussion see Soille[32].

Bibliography

[1] American Cancer Society. (2010). Breast Cancer Facts and Figures. Retrieved from http://www.cancer.org/acs/groups/content/@nho/documents/document/f861009final90809pdf.pdf

[2] Kumar, V., Abbas, A., Fausto, N., Mitchell, R. *Robbins Basic Pathology*, 8th Edition, Saunders, 2007.

[3] MacPhee, D. *Reproductive Histology*. Lecture Notes from MED-5600, Memorial University of Newfoundland, 2009.

[4] Rosen, P. *Breast Pathology*, 3rd Edition, Lippincott, Williams and Wilkins. 2009.

[5] Camp, R., Neumeister, V., Rimm, D. (2008). A Decade of Tissue Microarrays: Progress in the Discovery and Validation of Cancer Biomarkers. *Journal of Clinical Oncology*, 26(34): 5630-5636.

[6] Parker, R., Huntsman, D., Lesack, D., Cupples, J., Grant, D., Akbari, M., Gilks, C. (2002). Assessment of Interlabratory Variation in the Immunohistochemical Determination of Estrogen Receptor Status Using a Breast Cancer Tissue Microarray. *American Journal of Clinical Pathology*, 117: 723-728.

[7] Commission of Inquiry on Hormone Receptor Testing (Canada): http://www.cihrt.nl.ca

[8] Bacus, S., Flowers, J., Press, M., Bacus, J., McCarty, K. (1988). The Evaluations of Estrogen Receptor in Primary Breast Carcinoma by Computer-Assisted Image Analysis. *American Journal of Clinical Pathology*, 90(3): 233-239.

[9] Ficarra, E., Macii, E., De Micheli, G., Benini, L. (2006). Computer Aided evaluation of protein expression in pathological tissue images. *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems*.

[10] Kayser, G., Radziszowski, D., Bzdyl, P., Sommer, R., Kayser, K. (2006). Theory and Implementation of an Electronic Automated Measurement System for Images Obtained from Immunohistochemically Stained Slides. *Analytical and Quantitative Cytology and Histology*, 28: 27-38

[11] Lehr, H., Mankoff, D., Corwin, D., Santeusanio, G., Gown, A. (1997). Application of Photoshop-based Image Analysis to Quantification of Hormone Receptor Exprtession in Breast Cancer. *The Journal of Histochemistry and Cytochemistry*, 45(11): 1559-1565.

[12] Rexhepaj, E., Brennan, D., Holloway, P., Kay, E., McCann, A., Landberg, G., Duffy, M., Jirstrom, K., Gallagher, W. (2008). Novel Image Analysis Approach for Quantifying Expression of Nuclear Proteins Assessed by Immunohistochemistry: Application to Measurement of Oestrogen and Progesterone Receptor Levels in Breast Cancer. *Breast Cancer Research*, 10(5).

[13] Pham, N., Morrison, A., Schwock, J., Aviel-Ronen, S., Iakovlev, V., Tsao, M., Ho, J., Hedley, D. (2007). Quantitative image analysis of immunohistochemical stains using a CMYK color model. *Diagnostic Pathology*, 2(8).

[14] Schnorrenberg, F., Tsapatsoulis, N., Pattichis, C., Schizas, C., Kollias, S., Vassiliou, M., Adamou, A., Kyriacou, K. (January/February 2000) Improved Detection of Breast

Cancer Nuclei Using Modular Neural Networks. *Engineering in Medicine and Biology*, 48-63.

[15] Willemse, F., Nap, M., Henzen-Logmans, S., Eggink, H. (1994). Quantification of Area Percentage of Immunohistochemical Staining by True Color Image Analysis with Application of Fixed Thresholds. *Analytical and Quantitative Cytology and Histology*, 16(5): 357-364.

[16] Ruifrok, A., Johnston, D. (2001). Quantification of Histochemical Staining by Color Deconvolution. *Analytical and Quantitative Cytology and Histology*, 23: 291-299.

[17] Ruifrok, A., Katz, R., Johnston, S. (2003). Comparison of Quantification of Histochemical Staining By Hue-Saturation-Intensity (HSI) Transformation and Color-Deconvolution *Applied Immunohistochemistry and Molecular Morphology*, 11(1): 85-91.

[18] Brey, E., Lalani, Z., Johnston, C., Wong, M., McIntire, L., Duke, P., Partick, C. (2003). Automated Selection of DAB-labeled Tissue for Immunohistochemical Quantification. *Journal of Histochemistry and Cytochemistry*, 51(5): 575-584.

[19] Lehr, H., van der Loos, C., Teeling, P., Gown, A. (1999). Complete Chromogen Separation and Analysis in Double Immunohistochemical Stains Using Photoshop-based Image Analysis. *The Journal of Histochemistry and Cytology*, 47(1): 119-125.

[20] Malpica, N., de Solorzano, C., Vaquero, J., Santos, A., Vallcorba, I., Garcia-Sagredo, J., del Pozo, F. (1997). Applying Watershed Algorithms to the Segmentation of Clustered Nuclei. *Cytometry*, 28: 289-297.

[21] Bamford, P., Lovell, B. (1998). Unsupervised cell nucleus segmentation with active contours. *Signal Processing*, 71: 203-213.

[22] Mao, K., Zhao, P., Tan, P. (2006). Supervised Learning-Based Cell Image Segmentation for p53 Immunohistochemistry *IEEE Transactions on Biomedical Engineering*, 53(6):1153-1163.

[23] Schüpp, S., Elmoataz, A., Herlin, P., Bloyet, D. (2001). Mathematical Morphological Segmentation Dedicated to Quantitative Immunohistochemistry. *Analytical and Quantitative Cytology and Histology*, 23: 257-267.

[24] Latson, L., Sebek, B., Powell, K. (2003). Automated Cell Nuclear Segmentation in Color Images of Hematoxylin and Eosin-Stained Breast Biopsy. *Analytical and Quantitative Cytology and Histology*, 25: 321-331.

[25] Kostopoulos, S., Cavouras, D., Daskalakis, A., Bougioukos, P., Georgiadis, P., Kagadis, G., Kalatzis, I., Ravazoula, P., Nikiforidis, G. Colour-Texture based image analysis method for assessing the Hormone Receptors status in Breast tissue sections. in 2007 *Proceedings of IEEE conference on Engineering in Medicine and Biology.*

[26] Tuominen, V., Ruotoistenmaki, S., Viitanen, A., Jumppanen, M., Isola, J. (2010). ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Research*, 12: R56.

[27] Gokhale, S., Rosen, D., Sneige, N., Diaz, L., Resetkova, E., Sahin, A., Liu, J., Albarracin, C. (2007). Assessment of Two Automated Imaging Systems in Evaluating Estrogen Receptor Status in Breast Carcinoma. *Applied Immunohistochemistry and Molecular Morphology* 15(4): 451-455.

[28] Nassar, A., Cohen, C., Agersborg, S., Zhou, W., Lynch, K., Heyman, E., Olson, A., Lange, H., Siddiqui, M. (2011). A New Immunohistochemical ER/PR Image Analysis

System: A Multisite Performance. *Applied Immunohistochemistry and Molecular Morphology* 19(3): 195-202.

[29] Krecsak, L., Micsik, T., Kiszler, G., Krenacs, T., Szabo, D., Jonas, V., Csaszar, G., Czuni, L., Gurzo, P., Ficsor, L., Molnar, B. (2011). Technical note on the validation of a semi-automated image analysis software application for estrogen and progesterone receptor detection in breast cancner. *Diagnostic Pathology*, 6(6).

[30] Vrekoussis, T., Chaniotis, V., Navrozoglou, I., Dousias, V., Pavlakis, K., Stathopoulos, E., Zoras, O. (2009). Image Analysis of Breast Cancer Immunohistochemistry-stained Sections Using ImageJ: An RGB-based Model. *AntiCancer Research*, 29: 4995-4998.

[31] Prasad, K., Tiwari, A., Ilanthodi, S., Prabhu, G., Pai, M. (2011). Automation of immunohistochemical evaluation in breast cancer using image analysis. *World Journal of Clinical Oncology*, 2(4):187-194.

[32] Soille, P. *Morphological Image Analysis*, 2nd Edition, Springer, 2002.

[33] Otsu, N. (1979). A threshold selection method from grey-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9: 62-66.

[34] Vincent, L., Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6): 583-598.

[35] Vincent, L. (1993). Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms. *IEEE Transactions on Image Processing*, 2(2): 176-201.

[36] Berger, C., Koeijer, J., Glas, W., Madhuizen, H. (2005) Color Separation in Forensic Image Processing. *Journal of Forensic Science*, 51(1): 100-102.

[37] Duda, R., Hart, P., Stork, D. *Pattern Classification*, 2nd Edition, John Wiley & Sons, 2001.

[38] Keppel, G., Zedeck, S. *Data Analysis for Research Designs*, 11th Edition, W. H. Freeman and Company, 2006

[39] Rabinovich, A., Agarwal, S., Laris, C., Price, J., Belongie, S. (2003). Unsupervised Color Decomposition of Histologically Stained Tissue Samples. *Advancements in Neural Information Processing Systems*, Vancouver, BC, Canada.

[40] Lee, D.D., Seung, H.S., (1999). Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401: 788-791.

[41] Makretsov, N., Dawson, S.J., Mercer, J., Provenzano, E., Howat, W., Morris, L., Blows, F., Driver, K., Pharoah, P., Caldas, C. (2009). Automated Image Analysis of Nuclear Immunomarkers in Breast Cancer: Data on 2064 Cases Tissue Microarrays. Poster presented at the 2009 European Congress of Pathology, Florence, Italy. Available at `http://ponteiro.ath.cx/ecp2009eposters/posterFlash.asp?posterId=944&searchTe`.

[42] Gonzalez, R., Woods, R. *Digital Image Processing*, 3rd Edition, Prentice Hall, 2008.

[43] Diaz, L., Sahin, A., Sneige, N. (2004). Interobserver Agreement for Estrogen Receptor Immunohistochemical Analysis in Breast Cancer: A Comparison of Manual and Computer-Assisted Scoring Methods. *Annals of Diagnostic Pathology*, 8(1): 23-27.

[44] Turbin, D., Leung, S., Cheang, M., Kennecke, H., Montgomery, K., McKinney, S., Treaba, D., Boyd, N., Goldstein, L., Badve, S., Gown, A., van de Rijn, M., Nielson, T., Gilks, B., Huntsman, D. (2007). Automated quantitative analysis of estrogen receptor expression in breast carcinoma does not differ from expert pathologist scoring: a tissue microarray study of 3,484 cases. *Breast Cancer Research and Treatment*, 110: 417-426.

[45] Kong, H., Gurcan, M., Belkacem-Boussaid, K. Partitioning Histopathological Images: An Integrated Framework for Supervised Color-Texture Segmentation and Cell Splitting. *IEEE Transactions on Medical Imaging*. 2010.

[46] Ruifrok, A. (1997). Quantification if Immunohistochemical Staining by Color translation and automated thresholding. *Analytical and Quantitative Cytology and Histology*, 19: 107-113.